

Alma Mater Studiorum - Università di Bologna
in cotutela con University of Luxembourg - Université du Luxembourg

**DOTTORATO DI RICERCA IN
LAW, SCIENCE AND TECHNOLOGY**

Ciclo 35°

Settore Concorsuale: 12/G1 - DIRITTO PENALE

Settore Scientifico Disciplinare: IUS/17 - DIRITTO PENALE

**NEW PERSPECTIVES ON A.I. IN SENTENCING. HUMAN DECISION-MAKING
BETWEEN RISK ASSESSMENT TOOLS AND PROTECTION OF HUMAN RIGHTS**

Presentata da: Olimpia Barresi

Coordinatore Dottorato

Monica Palmirani

Supervisore

Vittorio Manes

Co-Supervisore

Silvia Allegrezza

Esame finale anno 2023



PhD-FDEF-2023-xxx
The Faculty of Law, Economics and Finance



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

The Department of Legal Studies

DISSERTATION

Defence held on 04-07-2023 in Bologna

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN DROIT

AND

*DOTTORESSA DI RICERCA IN LAW, SCIENCE AND
TECHNOLOGY*

by

BARRESI Olimpia

Born on 20-06-1994 in Palermo (PA) (Italy)

NEW PERSPECTIVES IN AI AND SENTENCING

Table of Contents

Abstract.....	8
A. Introduction to the research topic and research questions.....	10
B. Methodology	20
C. Territorial Scope: Italian and supranational landscape	22

Chapter One

1 Foreword: new players in the courtroom	23
1.1 The <i>boundaries</i> of the study: the continuation of the premise.....	29
2 Artificial intelligence: a brief outline of its origins and evolution.....	31
2.1 Machine learning, deep learning and dynamic risk analysis systems.....	35
3 Predictive analysis	41
3.1 The new players in analytical prediction: predictive algorithms	43
3.2 Brief outline of how algorithmic software works	45
4 The “shy entry” of Artificial Intelligence into the courtrooms	49
4.1 The first automated decisions in civil and administrative sectors: the disruptive technologies	53
4.1.1 Italian jurisprudence and the A.I.....	54
4.1.2 The Council of State's stance: the attempt to separate the concepts of algorithm and artificial intelligence	58
4.2 Predictive algorithms replace the judge: a look at the supranational landscape	59
4.2.1 The Estonian case and the algorithm solving low disputes	60
4.2.2 A look to the East: China and the new algorithmic prosecutor.....	62
5 The progressive approach to criminal justice in Italy	62
5.1 The three application scenarios: investigative, evidentiary and decisional	66
5.1.1 Brief remarks on possible applications in the investigative field	66
5.1.2 In the field of evidence: brief remarks	70
5.1.3 In the decision-making field: risk assessment tools.....	72
6 The spread of predictive justice	74
6.1 Predictability in jurisprudence as an incomparable value.....	76
7 The database system in maximising predictive justice and the calculability of judicial decisions.....	81
8 The fragmentation of the judicial decision: new needs in a justice system	85
9 Perspectives <i>de iure condendo</i> : towards predictive justice behind the scenes of the courtroom.....	87

10 Concluding remarks: the appeal of predictive algorithms and a transforming criminal justice system	89
---	----

Chapter Two

1 Foreword: a general overview on risk assessment tools	93
1.1 Brief historical reflection on origin in the US context.....	97
2 New challenges for the criminal justice process? Brief historical evolution of risk assessment tools	100
2.1 The four generations of risk assessment tools over time	102
3 The definition and operation of risk assessment tools	108
3.1 Criteria in the method of operation of risk assessment tools	111
3.2 Brief hints on risk factors in risk assessments	113
4 The combination of static and dynamic factors	115
4.1 A deeper look: the different types of risk in relation to factors	120
4.1.1 The clinical risk.....	120
4.1.2 The actuarial risk.....	121
4.1.3 The professional structured risk.....	123
5 A crucial step: risk identification	125
5.1 Static and dynamic risk factors	127
5.2 The dynamic risk factors.....	127
5.2.1 The criminogenic needs	127
5.2.2 Psychosocial needs.....	128
6 The I.N.U.S. conditions of the criminal behaviour	130
6.1 A case in point: the Risk Need Responsivity model.....	131
7 The dual application front of risk assessment tools in criminal justice	132
7.1 The application paradigm of risk factors	136
7.1.1 In an <i>ante delictum</i> application perspective	137
8 Brief notes on <i>post delictum</i> application uses (<i>Segue</i>).....	142
8.1 An overseas perspective: current applications of risk assessment tools in sentencing and recidivism risk assessment	142
8.1.1 The US model: between evidence-based practice and systems used.....	142
9 Other risk assessments used in the investigation phase	147
9.1 An example of risk assessment: the Public Safety Assessment (PSA).....	147
9.2 The PATTERN algorithm: the system to be taken as a model?	151
10 Brief remarks: the anticipation of an initial operating proposal (<i>Segue</i>)	152

10.1 The latest generation of risk assessment tools: why are they better than others in criminal risk assessment?.....	155
10.2 Overcoming the intelligibility of decision-making mechanisms	157
11 When risk assessment approaches Artificial Intelligence	159
11.1 Two characteristics compared: accuracy and predictive significance	159
11.1.1 The risks related to implicit bias	162
12 First Concluding Remarks on the Use of Risk Assessment in Criminal Justice: Towards Fair Treatment Justice?	164

Chapter Three

1 Foreword: the concept of social dangerousness.....	168
1.1 The debate between the classical school and the positive school	172
1.1.1 A closer look at the Positive School: the delinquent's dangerousness	176
2 The structure of social dangerousness in the Italian Criminal Code.....	178
2.1 The passage and journey of the birth of the prognosis	181
2.2 Current prognostic criteria in different cases and their limitations.....	182
3 The two different types of danger	183
3.1 When the legal system finds itself making prognostic assessments	184
3.2 A decision looking to the future: between limits and difficulties of prognostic evaluation.....	186
4 The judge's decision and the beyond reasonable doubt criterion in prognostic evaluations	187
5 Prognostic assessments referred to the judge in the Italian legal system.....	191
5.1 The different types of prognosis at the trial stages	193
5.2 Prognostic evaluation in security measures	194
5.3 Prognostic evaluation in prevention measures.....	195
6 The structural characteristics of the prognosis of dangerousness	196
7 How Artificial Intelligence intervenes in the judgement of dangerousness.....	199
8 The problem of defusing human cognitive bias: possible advantages in the use of predictive algorithms	202
8.1 The second step in risk assessment tools: the individual and the group	203
8.2 The advantages of a mixed algorithmic evaluation: the US example.....	204
9 Possible remedies: enhanced and explanatory justification of the new algorithmic indices.....	207

Chapter Four

1 Methodological premise: between the judge's decision and risk assessment tools.....	210
---	-----

1.1 The continuation of a premise: between the decision and the future with risk assessment tools	213
2 The most delicate phase left to the judge: the choice on the commensuration of the penalty.....	216
2.1 The criteria and the different types of penalty: how much the judge's discretion is gradually affected.....	221
2.1.1 Initial reflections on the Italian discretionary system	223
3 Focus: the application of the sentence and the judge's discretionary power	225
4 American judicial practice and the use of risk assessment: a special <i>focus</i> on the selective incapacitation movement theory and evidence-based sentencing.....	227
4.1 Segue: the penalty phase in the US system.....	231
4.2 How actuarial risk assessment came about	233
4.3 Compas: the Loomis case and the Wisconsin Supreme Court decision	235
4.4 The peculiarity of the decision: the decisional 'double phase' in the choice of penalty treatment.....	237
4.5 An application overview of the United States: the case of Virginia	240
5 From risk assessment in recidivism to sentence commensuration: why algorithms fit into sentencing	241
5.1 The intersection of two provisions at the stage of assessing the penalty treatment.....	242
6 Criminal discretion in the Italian legal system: the difficult framing	245
6.1 Sentence commensuration and criticised discretion	247
6.2 The 'capacity to commit offences' as an assessment that forces one to look into the future.....	250
6.3 The problem of prognostic evaluations.....	252
6.4 The answer to a question: why <i>prognosis</i> is considered so important in the choice of sanction treatment	253
7 Limits and differences in algorithmic evaluation in the Italian penal system.....	257
7.1 The limits posed by Article 220 of the Code of Criminal Procedure: is the principle in crisis?	260
8 Discretion and its combination of constrained and controlled in the criminal justice system.....	262
8.1 A new constrained discretion: the judge's free conviction in the face of new probative evidence. The Weakness of Articles 132 and 133 of the Criminal Code. The Praxeological Guidelines on Discretion.....	264
8.1.1 The problem of the weakness 'in the dark' of prognostic judgements	268
8.2 The concept which returns: social dangerousness within Article 133 of the criminal code.....	271
8.3 Problems and the first emerging evidence on the phenomenological level.....	273
9 The human decision and the technological decision: a surmountable opacity? An adversarial 'technicalisation'	275
9.1 The algorithm in the decision-making phase: what benefits and towards	

what future?	277
10 The paradigmatic value of Article 133 of the Criminal Code: inadequate <i>criteria</i> ?	278
10.1 The <i>ethicality</i> of human judgement and its ineradicable subjective components	280
11 Concluding remarks: drawing conclusions on risk assessment	281

Chapter Five

1 Methodological premise: towards the proposition of a model and a look at the past	283
2 Future perspective: the proposed model	286
2.1 The output of the algorithm: predictions.....	289
2.1.1 At which stage could predictions intervene?	291
3 An initial critique of accuracy and the risk of generalisations	294
4 Risk indicators: the difficulty of selection	297
4.1 The most sensitive issue: a chance to overcome. The choice of risk factors	300
5 The forward-looking perspective and its structure: individualised judgement and human control of the judge	302
5.1 A <i>key</i> to begin with: the algorithm applied only <i>in bonam partem</i>	305
6 Comparing rights: the feasibility of a proposal between a balance of rights and guarantees	307
6.1 The narrower frame of applicability: the possibility of intervention in the face of relevant factors with dynamic characteristics	310
7 The main issues arising from the first reflections	311
7.1 Constitutional limits.....	312
7.2 Compression of personal liberty: between rights and balances under Article 5 ECHR.....	313
7.3 The risk of profiling: the use of big data and the invasive approach with individuals. Article 22 GDPR and its regulatory boundaries	314
7.3.1 The collateral risks associated with profiling: the delicate drifts of stereotyping.....	316
8 The problem of discrimination overcome by the rationality of the machine?	321
8.1 The problem of data transparency and the opacity of A.I. systems	322
8.1.1 Current <i>scenario</i> and possible future solutions.....	325
8.1.2 The risk of undermining the guarantee of the 'equality of arms at trial': the right of access for the defendant	326
8.1.3 A possible solution to overcome the obstacle of obscurity: towards greater transparency	330

9 The necessity and clash of penal guarantees	333
9.1 The difficult balance between presumption of harmlessness and presumption of innocence	333
9.1.1 The risk of determinism in decisions	335
10 The regulatory framework: the regulatory sources of Artificial Intelligence	337
11 New perspectives and positions on artificial intelligence: The EU White Paper	338
12 The Ethic Charter of EU	339
13 The Council of Europe's position on automated decisions with profiling	342
14 Regulatory limits to artificial intelligence: ethical and legal barriers?	343
14.1 The General Data Protection Regulation	346
14.2 Automated data processing: a step forward to the GDPR with Legislative Decree No. 51 of 2018	347
14.3 EU Directive 680/2016 on the processing of personal data for the prevention, investigation, detection and prosecution of criminal offences or the execution of criminal penalties	348
15 The proposal for a future regulation on A.I. <i>de iure condendo</i> perspectives. Progress towards regulation?	350
Conclusions.....	353
Bibliography	360
Books	360
Chapters in Edited Books.....	367
Articles.....	371
Online sources.....	390
Reports	393
Papers and Conference Presentations	394
Doctoral Theses	394

Abstract

The aim of this thesis is to investigate a field that until a few years ago was foreign to and distant from the penal system. The purpose of this undertaking is to account for the role that technology could play in the Italian Criminal Law system.

More specifically, this thesis attempts to scrutinize a very intricate phase of adjudication. After deciding on the *type* of an individual's liability, a judge must decide on the *severity* of the penalty. This type of decision implies a prognostic assessment that looks to the future. It is precisely in this field and in prognostic assessments that, as has already been anticipated in the United States, instruments and processes are inserted in the pre-trial but also in the decision-making phase. In this contribution, we attempt to describe the current state of this field, trying, as a matter of method, to select the most relevant or most used tools. Using comparative and qualitative methods, the uses of some of these instruments in the supranational legal system are analyzed. We do so to better enable policy makers and academics to understand the nuance that might arise from the introduction of such instruments, trying to abandon an approach of total closure and caution, but trying to glimpse and focus the analysis on certain instruments that might prove useful at a limited stage of the decision.

This approach makes it possible to take a closer look at the impacts on criminal and procedural rights and guarantees and the benefits they could provide.

Focusing attention on the Italian system, an attempt was made to investigate the nature of the element of an individual's 'social dangerousness' (*pericolosità sociale*) and capacity to commit offences, types of assessments that are fundamental in our system because they are part of various types of decisions, including the choice of the best sanctioning treatment. It was decided to turn our attention to this latter field because it is believed that the judge does not always have the time, the means and the ability to assess all the elements of a subject and identify the best 'individualizing' treatment in order to fully realize the function of Article 27, paragraph 3 of the Constitution.

Clearly, it has been acknowledged that the introduction of such instruments must necessarily be confronted with a system of substantive and procedural guarantees that must remodel or innovate in the presence of such instruments. Indeed, in a procedural dialectic that has hitherto seen only the defendant, the judge and other marginal subjects as protagonists, the I.A. instruments are part of this dialogue.

An attempt has been made in this paper to show an optimistic outlook towards an introduction of such tools, albeit only limited to a certain point in time and initially applied

only *in bonam partem*. The proposal considers ad hoc regulation of artificial intelligence in the first place and of such instruments thereafter as necessary. Reflections can at present only stop at an embryonic stage, since it must first be understood through a more in-depth study whether it is really acceptable to link the quantification of the sentence to the assessment of the risk of reoffending. We ask to what extent and in what ways some A.I. tools may prove not only useful but also crucial in a decision as support to the adjudicating body.

The significance of this study is that it informs our theoretical understanding of the relation between A.I. tools and criminal sectors by introducing a focus on risk assessment tools in the assessment of social dangerousness in the choice of the best correctional treatment. These findings indicate the need for resources in order to investigate this field.

A. Introduction to the research topic and research questions

In recent years, Artificial Intelligence¹ has been making an impact in many areas of society. While the technology (also referred to as A.I.) is already present in many aspects of our daily lives, there has recently been talk about Artificial Intelligence tools that can also be applied in the field of criminal justice².

Upon closer inspection, Artificial Intelligence represents a set of processes in which a series of tools possess characteristics that make them *autonomous* in performing certain tasks. Although until some time ago it would have seemed unthinkable to bring such advanced technology in contact with criminal justice, today the *landscape* has changed, and criminal lawyers and scholars are forced to deal with new subjects that possess more technical and 'less human' characteristics. All these elements, however, must always be confronted with instruments that are created by man and that therefore 'work' and 'process' data that are first entered and selected by human beings. The research aims to provide an answer to the main research question:

What role can A.I. technologies play within the criminal justice process? Is it acceptable to link the quantification of the sentencing to the assessment of the risk of recidivism? How accurate and reliable are the premises and results of this risk assessment?

This first central question starts from the assumption that the decision-making phase left entirely to the judge implies a series of evaluations and judgements partly linked to an 'absolute discretion', especially when it comes to evaluations concerning future probabilities. Starting from this assumption and from the complexity of certain types of evaluations, to which one must add the need and the community's demand for certain kind of punishment, it is an open question whether technologies can fit into a system that was born human and that itself has within it a series of human implications that are perhaps difficult to give up.

¹ A.I. has recently been defined as «software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal». The definition was given by the High-level expert group on Artificial Intelligence, A definition of AI: main capabilities and disciplines, April 2019.

² In some countries, the use of these instruments is becoming more widespread in sentencing; in fact, they are already used to solve low-value disputes. In Estonia, for example, low-value civil disputes of EUR 7,000 have been automated and are resolved by algorithms.

Firstly, it should be noted that the areas within which the technological revolution set in motion by A.I. could more significantly impact on the claims of protection of the legal matters, entrusted to the criminal law, are basically four: the activities of law enforcement and, in particular, predictive policing, where A.I. systems can provide an important contribution to counter, or better still, prevent, the commission of crimes; the possible use of decision-making algorithms to solve criminal disputes, so as to operate as a sort of replacement, or at least side-by-side, of the judge-man with the judge-machine; the evaluation of the criminal danger entrusted to predictive algorithms, able to draw on and re-elaborate enormous quantities of data in order to bring out relations, coincidences, correlations, which allow the profiling of a person and prediction of his behaviour, also of criminal relevance; finally, the possibility of the involvement - as instrument, as author or as victim - of an A.I. system in the commission of a crime³.

More specifically, it is possible to divide the introduction of these technologies into the criminal justice system into two directions where it is more widespread and debated: in terms of *prevention*, as tools used by police forces to prevent and improve the use of resources in crime detection and prevention, and in *pre-trial decision-making and sentencing*; this latter field will be the focus of this research.

When Artificial Intelligence approaches criminal law, there are several areas and institutions with which it intersects: the issues of liability and causality that have emerged following the spread of driverless cars⁴, or the liability profiles that emerge as result of machine error⁵, or even the ethical-philosophical issues related to the “thinking” and the humanity of the algorithm⁶.

On closer inspection, the interest in this topic arose precisely from the question of what is meant by *human judgement*. The questions and issues emerge even more when one thinks of the judging body that finds itself having to make a judgement not only on the guilt of an individual for a given fact, but also and above all when the same is required to make a judgement that also go beyond the objective data (if one can define them in this way) and more

³ F. BASILE, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Diritto penale e uomo*, and also in *Diritto penale contemporaneo*, 29 September 2019, 1 ss.

⁴ To date, within the Council of Europe, only a limited number of countries have adopted general regulations for the use of automated driving (Austria, Germany, France and Switzerland), using the traditional concepts of the various liability schemes, while other countries have only adopted specific regulations on pilot tests.

⁵ See the Report from the Commission to the European Parliament, the Council and the European economic and social Committee. Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics, Brussels, 19 February 2020.

⁶ On an interesting read on the 'ethics of algorithms', see B. D. MITTELSTADT - P. ALLO M. TADDEO - S. WACHTER - L. FLORIDI, *The ethics of algorithms: Mapping the debate*, in *Big data society & press*, October 2016, 4.

objective in its possession (all the elements pertaining to the fact of the crime and the circumstances and all the elements), but above all in all those cases in which the judge finds himself having to deal with a prognostic assessment, that is one that looks to the future. Reference is made to all those cases in which the judge must assess, on the basis of the elements at his disposal, what may be the degree of risk of recidivism or of a capacity to commit offences for the purpose of deciding on a possible application of a precautionary or preventive measure, or even in the *sentencing phase*, of commensuration and choice of punitive treatment.

As will be analysed during the development of the entire work, there are several cases in which the judge, in the Italian legal system, finds himself having to make various predictive judgements on the dangerousness of the defendant. Indeed, among the most delicate of course are those in the field of the application of precautionary measures, relating to the existence of the *periculum libertatis*, as well as, and this is the aspect that is of most interest here, in the phase of determining the penalty on the offender's capacity to commit a crime (a concept that is very vague and difficult to delineate its contours both at the level of demonstration and at the level of judgement)⁷.

The work critically analyses the perspectives linked to the use of predictive algorithms to assess social danger, in the awareness of the advantages that may derive from an evidence-based ascertainment system, on the one hand, and of the risks for the protection of fundamental rights, on the other; can the algorithm represent an expert opinion, submitted to the judge's scrutiny by adopting the 'Daubert' criteria?

Therefore, an attempt will be made to follow three macro-directives during the development of the thesis (which will be further defined below).

The thesis is divided into five chapters.

In the first part, an attempt will be made to illustrate the state of the art of these instruments and to verify the validity of the model (even if this requires, with a view to the future, a more technical analysis that calls into play other legal practitioners and others); secondly, in proposing these instruments, the data and thus their ingenuity and reliability will play a fundamental role; only in a third instance will an attempt be made to assess and verify the compatibility of this possible applicative proposal within the constitutional limits and spaces and the regulations set up to protect them at a supranational level.

⁷ That this is to be assessed as a prognosis of the agent's future conduct follows from a constitutionally oriented reading of the provision of Article 133(2) of the criminal code. See for all, G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di Diritto Penale. Parte generale*, 7a ed., Milan, 2017, 706.

Having broadly outlined the guidelines followed during the study of these years, the analysis of the paper will then focus precisely on this aspect.

In the first chapter, an attempt will be made to provide an answer to the following *subquestions*:

- What is Artificial Intelligence?
- What are the tools of Artificial Intelligence that fit into criminal law?
- Which States are already applying them? Under what conditions can they be applied?

In the first chapter, an attempt will be made to frame the state of the art of the subject, the reasons for the approach of these instruments to criminal justice, the first decisions that have emerged in other areas of law in Italy and abroad.

In the remainder of the paper, an attempt will be made to provide an overview of the main Artificial Intelligence tools already in use today in certain phases of criminal justice in the United States and the tools that, due to their characteristics, may be of assistance to legal practitioners. To this end, an attempt will be made to trace a path of investigation that will be aimed at investigating the static and dynamic aspects of the new actors at play. We will approach the heart of the analysis carried out in the paper and provide an overview of the key players in the paper, namely the risk assessment tools (tools already used in various sectors), the functioning and structure of which are considered suitable for carrying out these types of assessments.

From a more practical point of view, an attempt will then be made to question the possible dual function of these instruments in the Italian legal system.

In the second chapter, risk assessment tools will be introduced. An initial descriptive and qualitative analysis of them will be attempted and the following *subquestions* will be answered:

- What are risk assessment tools in criminal law?
- Which risk assessment tools are most widely used in the supranational scene today?
- What are the main characteristics of the actuarial method and why is it so important in assessing the dangerousness of the individual?

Indeed, the paper will go on to examine a second spectrum of application: assessing how these instruments can also be used to identify the best sanctioning treatment, in a cost-benefit

analysis that will see as its final result the choice of the best treatment⁸, with a view to a maximum *re-valorisation* of the re-educative principle of punishment pursuant to Article 27, paragraph 3 of the Constitution. Lastly, the difficult balance and equilibrium between new emerging rights and ancient constitutional and penal guarantees will be analysed.

Undoubtedly, in a *de iure condendo* perspective, consideration will be given to the upheavals and new arrangements that may arise if the introduction of such instruments is admitted. Consideration will be given to the idea of a changing judgement, to its essential characteristics and to how much, we already anticipate, the idea of the human judge is *necessarily inalienable*.⁹

Subsequently, in the third and fourth chapters, we will first trace the characteristics of the concept of *social dangerousness* (*'pericolosità sociale'*) in Italy and its evolutions and declinations which, over time, have contributed to delineate it as we know it today; furthermore, we will try to evaluate how we can admit the change of a sentencing that accompanies the human judge's evaluation with a support tool that is able to help the latter in his final choice. Such a tool is believed to be able to accompany the judging body in making certain evaluations in which it is necessary to be aware of all the elements that concern a subject and, in a second moment, to support it in an evaluation that looks to the future, being fully aware that the sanctioning treatment and the choice of the same, cannot be limited to a present moment and even 'retrospective' but must necessarily look to the future.

In fact, an attempt will be made in these two chapters to answer the following *subquestions*:

- What is the aim of incorporating such tools into the criminal justice process?
- What does it mean to "*decide the treatment for a judge*"? (

⁸ For an approach that examines the hermeneutic interpretation of the judge and the overcoming of the axiom of 'more probable than not' see on the topic, M. CATERINI, *Il giudice penale robot*, in *Giustizia penale e nuove tecnologie*, 19th December 2020.

⁹ M. TARUFFO, *Judicial Decision and Artificial Intelligence*, in *Artificial Intelligence and Law*, 1998, 316-317, reiterated the enormous difficulty and essential characteristics of legal reasoning that can hardly fit into predefined models, saying "If one considers the evident feature of complexity, variability, flexibility and discretion that are typical of judicial decisions, any approach aimed at interpreting the judicial reasoning according to logical rules and models may appear as doomed to failure. In fact, the history of the logical theories of judicial reasoning is largely a history of misunderstandings, errors, manipulations and defeats. [...] On the one hand, one may observe that the main attempts to "computerize" the reasoning of the judge were so rough, and unable to interpret the complex nature of decision-making, that they could not succeed in producing reliable models of the judge's reasoning. These attempts, one might add, are a good proof of the impossibility of interpreting such a reasoning in terms of A.I. On the other hand, one may consider that the decision-making procedure is so complex, variable, uncertain, fuzzy, and value-laden, that it could never be reduced to logical models. Any logical model, one might say would necessarily leave aside important features of the decision-making reasoning that cannot be reduced to logical forms. Therefore, such a model would be basically false a description and inappropriate as a prescriptive model for judges".

- What criteria and parameters does a judge use when assessing the commensuration of a penalty?
 - What are prognostic evaluations? What elements are they composed of?
 - What are the first effects and implications of these first supranational applications?
 - What decisional 'discretion' means for a judge in the commensuration of punishment?
 - Would it be necessary to *readjust* conditional discretion to a new type of discretion that takes into account algorithmic results?
 - Is incapacitation, resulting from the extension of the indeterminate sentence, the only possible way to address the social dangerousness of the offender? Why should the legislature entrust the assessment of a defendant's dangerousness to an algorithmic system?

The last chapter will focus on providing an overview of national and supranational legislation surrounding the possible and future introduction of such instruments, with a view to balancing new rights and guarantees.

In particular, the last chapter will also focus on 'drawing conclusions' on the applications described above and the possible proposal, trying to glimpse and mark the existing regulatory limits and boundaries. For this reason, an attempt will be made to answer the following sub-questions in the conclusion:

- What are the prospects of impartiality and accuracy that an output provides us in 'predicting' the risk of reoffending? And what limits can criminal law place on the technological evolution represented by risk assessment tools based on Artificial Intelligence?
 - What Artificial Intelligence tools exist to process data and related risks?
 - What issues arise from the use of *big data*?
 - It would be possible to establish a system of open data in the judicial system, in order to respect the protection of data (GDPR regulation)?
 - What kind of ethical and legal issues arise in the idea of judge and machine helper?
 - What is the current EU legal framework in this field?
 - What kind of regulation is there today on the topic of A.I. and criminal law?

- What are the roots and reasons behind the need to introduce these new technologies into the process?
- What is meant by “*fair justice*”?

The purpose of this research will be to provide a framework and a proposal on how specific Artificial Intelligence tools can help improve certain aspects of criminal justice. More specifically, the analysis and the study will focus on the possibility of providing tools to judges in the evaluation and commensuration of punishment and for introducing into Italy's criminal law judicial system the practice of risk assessment (R.A.)¹⁰ of criminal recidivism and future violence.

The interest in this topic certainly stems from the recent effects of the digital revolution on a traditional field of scientific reflection and empirical research, namely that of the determination of social dangerousness and its inclusion among the instruments of criminal law. This is a field of research on the borderline between criminal law, the philosophy of law and psycho-criminology, which has always pushed towards the creation of quantitative risk models (of recidivism and violent behavior), to which the most recent and effective computational resources can now be usefully applied. It is essential, however, to retrace the evolutionary lines of the phenomenon in order to prevent it from ending up shrouded in the ambivalent attitude – of both fascination and dystopia – that today surrounds computational sciences and, in particular, artificial intelligence.

The use of A.I. science and technology in criminal matters poses specific challenges as its application may reflect some current public debates about the alleged predictability of offending behavior and about the possibility of introducing them in the sentencing phase¹¹.

A final concluding remark is directly related to the peculiarity of the decision-making system in Italy; in fact, in the Italian legal system, the sentencing phase is linked to and characterized by very precise indexes which the judge is required to follow and is surmounted by a series of inalienable guarantees. In Italy there is a type of discretion given to judges which is “constrained”. And the instrument of critical control of this constrained discretion is the

¹⁰ Risk assessment can provide an empirical estimate of whether an offender has a sufficiently high likelihood of again committing crime to justify incapacitation. That is, within a range of severity set by moral concerns about the criminal act of which the offender has been convicted, risk assessment can assist in determining whether, on utilitarian crime control grounds, an offender should be sentenced to the upper-bound of that range. See J. MONAHAN - J. L. SKEEM, *Risk assessment in criminal sentencing*, in *Annual Review of clinical psychology*, December 2015, 493 s.

¹¹ “Criminal sentencing is one of the most difficult responsibilities of judging”, see M. E. Donohue, *A replacement for justitia's scales? Machine learning's role in sentencing*, in *Harvard Journal of Law & Technology*, Vol. 32, No. 2, Berlin, 2019; E. MARVINE - FRANKEL, *Criminal sentences: law without order*, 15–16, 1972.

statement of reasons, at the center of which are the reasonableness and validity of the arguments put forward in support of it, or in support of the choices made in relation to the commensuration of the penalty (art. 133 Criminal Code)¹².

However, this reflection stems from the premise that it is considered that the parameters provided to the judge are denoted by a wide flexibility that leaves enormous discretion to judges. In fact, in the face of a system that shows gaps as regards "excessive discretion"¹³, to cope with and to reduce or eliminate "the uncertainty of the decision", it is considered useful to investigate this field in order to provide a picture of the matter and possible proposals. The former is aimed at punishing the offence considered from a purely objective point of view, the latter, on the other hand, is designed to prevent the offender's unlawful conduct originating from a judgment of social dangerousness made on his personality. While punishment presupposes guilt and therefore all the subjective elements on which criminal responsibility is based, the security measure presupposes social dangerousness, i.e. a prognostic judgement made on the subject that would allow understanding whether, in the future, he is capable of committing other crimes. It is precisely based on this premise that the need to establish precise criteria and parameters on which the judge must rely in exercising his discretionary activity is founded. There is a specific provision in the Italian Criminal Code concerning the concept of 'social dangerousness'. In fact, Article 203 of the Italian Criminal Code states: "For the purposes of the criminal law, a person is socially dangerous, even if he cannot be charged or is not punishable, who has committed any of the facts indicated in the previous article, when it is probable that he will commit new facts foreseen by the law as crimes. The quality of a socially dangerous person is inferred from the circumstances indicated in Article 133 c.p.”.

Faced with a *landscape* which is constantly and incessantly evolving, today's jurist is called upon to question and answer the new questions; it is a task that must be undertaken wisely by weighing up the interests at stake and, at the same time, with balance; it is a task that he cannot

¹² V. MANES, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in *Discrimen*, 15 May 2020, 2.

¹³ Among those who argued in favour of the introduction of such tools as a support for human decisions, see V. CHIAO, *Fairness, accountability and transparency: notes on algorithmic decision-making in criminal justice*, Cambridge, 20 June 2018, "In contexts where decisions are left to the relatively unstructured discretion of a human decision-maker, there is some sense to providing an opportunity for adversarial disputation. However, we should not assume that adversarial disputation will continue to be equally valuable in contexts where predictive algorithms turn out to be substantially more reliable than human decision-makers. In those contexts, public accountability, in the sense of ensuring that decisions are as likely to be correct as we can manage, is probably not best fostered by having individual accused challenge the technical details of an algorithm in the course of their criminal proceedings”.

shirk, on pain of an anachronistic rearguard position that would, however, soon risk being overtaken by the impetuous advance of the process of digital transformation of our society¹⁴.

It would indeed seem to propose a dystopian *landscape*, where human judgement is 'artificially' rendered less human. However, this is not the aim of this research, which instead proposes to present the current panorama of A.I. tools that could approach criminal justice in Italy as well, trying to provide an analysis that looks not only at the benefits, but also at the risks mostly related to the possible violations of constitutionally protected rights and the guarantees safeguarded in the penal code and criminal procedure code. Questions concerning the artificial intelligence-justice pair inevitably have repercussions on the figure who holds the decision-making power in the process and who is the interpreter of justice itself.

The reflection on the role of the judge and on the advisability of him being outflanked – or even replaced – by the decision-making machine then becomes fertile ground for confrontation and the search for new balances, in the complex dialogue between law and technology. It is in this perspective, therefore, that we shall analyse the merits and demerits of the new 'algorithmic justice'¹⁵.

In conclusion, the motivations that pushed me towards an in-depth study of this topic stemmed from the enormous fascination aroused by these instruments and the numerous questions that automatically arose upon seeing the first applications in the US legal system. Moreover, the first reflections and curiosity focused on understanding how in reality there are many cases in which the judge is confronted with decisions that oblige him to make assessments that look to the future. Therefore, as will be seen during the research, these are evaluations that present a degree of uncertainty and an extra effort on the part of the judging body, which finds itself having to analyse and take into consideration all the elements in order to then be able to choose the best sanctioning treatment for the individual.

Therefore, the thesis aims to propose new keys to interpretation and to lay the foundations for a future investigation that would look at such tools not only as enemies of criminal law guarantees but as tools from which beneficial use could be made for the improvement of a typology of complex evaluations. All this does not require attention to a necessary regulation of artificial intelligence that does not yet exist and, at the same time, understanding how these

¹⁴ P. SEVERINO, *Intelligenza artificiale e diritto penale*, in U. Ruffolo (ed), *Il diritto, i diritti, l'etica*, Milan, 2020.

¹⁵ The application aspects of digital justice are also dealt with in depth by R. BICHI, *Intelligenza artificiale, giurimetria, giustizia predittiva e algoritmo decisorio. Machina sapiens e il controllo sulla giurisdizione*, in U. Ruffolo (ed), *Intelligenza artificiale*, 423-447.

same tools can fit within the criminal justice system while respecting the individual's guarantees and rights.

B. Methodology

During the course of the thesis, an attempt is made to answer the main research questions as outlined above. Indeed, the methodology used in this research will combine the study of literature, legislation and underlying principles and case law. In particular, it will combine and intersect with a guarantor and legal approach based on the study and foundations of criminal law, which will be combined with the presentation of risk assessment tools adapted to practical and legal use. In the background there will always be a reflection that looks at the comparison between the institutions and legislation of other systems. In particular, the comparative analysis will focus mainly on American countries on which it is possible, through a critical and analytical method, in order to assess what the first impacts are following the first applications of these tools. The analysis will therefore be carried out in an interdisciplinary manner of the two relevant subjects.

The documentation collected for the elaboration and drafting of the aforementioned work is largely made up of wide-ranging academic texts and articles ranging from Anglophone, French and Italian literature. The choice made is undoubtedly dictated by the desire to make the state of the art in the field of artificial intelligence and criminal justice as broad and clear as possible. Moreover, documents are a good source of qualitative data; they include different types such as: bibliographic sources, newspaper articles, websites.

With regard to the structural layout of the first chapters, it was deemed appropriate to use the classic tool based on the collection of material found on the web, international articles, monographs and texts that could offer a clear and diachronic picture of geographical developments regarding tourism. The bibliography of the Italian matrix was obtained through bibliographic research using the *Sebina* system: for the most part these are texts found within the libraries of Bologna's circuit, where the research was compiled and therefore found through inter-library loan. This initial collection, which, as it should be noted, is not assumed to be exhaustive but as valid material and as an additional piece to the studies carried out on the subject, has thus constituted a load-bearing base on which to base the entire corpus of the secondary phase, the one that is clearly indicated as the most applied part.

In view of what has been said so far, for systematic reasons, we intend to focus the analysis on only a few profiles pertaining to the application of algorithmic tools to criminal proceedings. Since it is not possible here to deal in sufficient depth with all the various aspects connected with the interaction between A.I. and criminal law, we will omit to examine many issues directly connected with criminal law institutions (such as *guilt*, *causality*, *liability*). Rather, in

keeping with the *focus* of the paper, the study will concentrate on the interference of intelligent systems in the phases that make up the criminal judgement, flanking an initial analysis of the various stages that precede the exercise of punitive action with some considerations on the usefulness of AI and algorithms at the moment of interpretation and application of the law, as well as in the commensuration of the punitive treatment. To this end, we intend to conduct the research from a comparative perspective, to place the Italian case in the European context as well as with respect to the most advanced international scenarios.

This part of the research avails itself of the use of different tools and methodologies than those used previously, which concern not only a bibliographic collection, but also a change of method that complements the knowledge already acquired and consolidated to a greater extent during the three years of doctoral study, which allowed for an appropriation with more applied implications. The digital tools presented within this work allow for a parallel narration of the topics surveyed through multimedia support.

This method of research was chosen because it is considered the most appropriate method for conducting this type of study, which is still in its embryonic form, being able, for the time being, to stop at embryonic considerations on the first practical implications of these instruments in the supranational field. It was possible to conduct this study, enriching it with copious participation in seminars, meetings, readings on the subject, and numerous conferences held both in Bologna and Luxembourg during the period of study in *joint-supervision*.

In conclusion, using a qualitative method and approach, an attempt has been made to approach the topic by following foreign and national literature. There will also be ethical considerations on the essence and rationality of human judgement.

C. Territorial Scope: Italian and supranational landscape

Another fundamental aspect of this study, which must be framed from the outset, concerns the more purely geographical aspect.

Clearly, for the sake of method and conciseness, we have chosen to focus the analysis and possible introductory proposal only on the Italian legal system.

It is not a choice that derives, in fact, only from analytical reasons, but rather because it is considered extremely necessary to start the reflection from a specific type of evaluation - which does not belong only to the Italian legal system - but which is nevertheless found in multiple forms in the domestic legal system. In fact, as will be seen later in the paper, during criminal proceedings, the Italian judge is called upon to make various predictive judgments on the dangerousness of the defendant (from precautionary measures to the prognostic judgments already inherent in the assessment and choice of punitive treatment).

We shall not fail, however, especially in the final part of the paper, to recognize the contours and characteristics of the regulatory framework at the national and supranational level. Having overcome this first and more general premise, it is immediately anticipated that the theme that occupies the study interest of this thesis belongs to a domain that is still considerably influenced by the differences between the American and continental European approaches and thus, on a more strictly legal level, by the macro distinction between common law and civil law. Indeed, as will be seen in the course of the development of the paper, the entire debate on the functions of criminal law and the various theories of punishment is strongly drawn from the historical background and evolution of the individual legal systems, with a (since time immemorial) permanent and significant caesura between the two predominant large families: that of the Romano-Germanic tradition belonging to continental Europe and, on the other hand, the common law tradition.

For these reasons, the ultimate aim of this work is to draw, at first, a picture of the current state and the strong topicality of the discussion on the role that crime prevention can play in the quantification and commensuration of punishment and, at a second time, starting from the analysis and historical evolution of the concept of dangerousness, to try to, after taking into consideration the traditional theories on risk assessment, how they can be reconciled with the resources and new technologies of A.I., computational sciences and, *last but not least*, with the system of fundamental rights that permeates all legal systems and, specifically, the Italian one.

Chapter One

Artificial Intelligence in contact with justice

The current state and first considerations

SUMMARY: 1. Foreword: new players in the courtroom – 1.1. The boundaries of the study: the continuation of a premise – 2. Artificial intelligence: a brief outline of origins and evolution. – 2.1. Machine learning, deep learning and dynamic risk analysis systems. – 3. Predictive analysis. – 3.1. The new players in analytical prediction: predictive algorithms. – 3.2. Brief outline of how algorithmic software works. – 4. The “shy entry” of Artificial Intelligence into the courtrooms. – 4.1. The first automated decisions in civil and administrative sectors: the disruptive technologies. – 4.1.1. Italian jurisprudence and the A.I. – 4.1.2. The Council of State's stance: the attempt to separate the concepts of algorithm and artificial intelligence. – 4.2. Predictive algorithms replace the judge: a look at the supranational landscape. – 4.2.1. The Estonian case and the algorithm solving law disputes – 4.2.2 A look to the East: China and the new algorithmic prosecutor. – 5. The progressive approach to criminal justice in Italy. – 5.1. The three application scenarios: investigative, evidentiary and decisional. – 5.1.1. Brief remarks on possible applications in the investigative field. – 5.1.2. In the field of evidence: brief remarks. – 5.1.3. In the decision-making field: risk assessment tools. – 6. The spread of predictive justice. – 6.1. Predictability in jurisprudence as an incomparable value. – 7. The database system in maximizing predictive justice and the calculability of judicial decisions. – 8. The fragmentation of the judicial decision: new needs in a justice system. – 9. Perspectives *de iure condendo*: towards predictive justice becoming real. – 10. Concluding remarks: the appeal of predictive algorithms and a transforming criminal justice system.

1 Foreword: new players in the courtroom

Parallel to the development of new technologies, Artificial Intelligence (A.I.)¹⁶ is becoming part of several and varied aspects of everyday life¹⁷. Indeed, there are many fields in which it is entering, simplifying many human activities and making them more usable for everyone.

On closer inspection, the scientific evolution, especially in recent years, has greatly enhanced the capacity for analysis, knowledge and development of various activities and phenomena in the daily life of people¹⁸. In particular, as Stephen Hawking already anticipated

¹⁶ The literature on the subject of Artificial Intelligence (A.I.) has, to date, been vast. For a general overview, in this first part, we will merely recall, D. HEAVEN (ed), *Macchine che pensano. La nuova era dell'intelligenza artificiale*, Bari, 2018; G.F. ITALIANO, *Intelligenza artificiale: passato, presente, futuro*, in F. Pizzetti, (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Turin, 2018, 216; J. KAPLAN, *Intelligenza artificiale. Guida al futuro prossimo*, Rome, 2018; A. VESPIGNANI, *L'algoritmo e l'oracolo*, Milan, 2019.

¹⁷ As has also been said, AI 'is everywhere', M.A. BODEN, *Intelligenza artificiale*, in J. I-Khalili (ed), *Il futuro che verrà*, Turin, 2018, 133.

¹⁸ D. POLIDORO, *Tecnologie informatiche e procedimento penale: la giustizia penale “messa alla prova” dell'Intelligenza artificiale*, in *Arch. Pen.*, No. 3, 1.

a few years ago, sooner or later the current productive world, as we know it, 'could be subverted by a particular form of artificial intelligence'¹⁹.

Today, it is possible to witness of a new 'technological wave'²⁰, a phenomenon of advancing technology that infiltrates various sectors and possesses unlimited expansion capable of affecting individuals and the community.

In this regard, Artificial Intelligence has made its entrance in recent years into several areas of justice, particularly in the civil and administrative fields. It does not seem difficult to grasp the strength of certain tools capable of making these areas more efficient, as the perspectives that are proposed and found in the background in this type of application show a landscape with fewer implications and application conflicts than in areas of justice such as criminal law²¹.

It is, therefore, important to anticipate that the enormous fascination with Artificial Intelligence has ended up persuading and making its way into the field of criminal justice²² as well, increasing from various points of view, the typology of evidentiary tools on the part of the trial parties, the recognition and search systems of individuals, and has ended up, in part, also modifying that aspect of the actual judgement placed in the hands of the judge²³.

Nonetheless, it is already worth noting how the panorama that lies ahead sees Europe being more "cautious" with respect to automated decision-making tools²⁴ than North America; it is

¹⁹ These statements and her position she expressed at a web summit in 2017. For further details, see veda http://www.repubblica.it/scienze/2017/11/07/news/stephen_hawking_1_intelligenza_artificiale_potrebbe_distribuire_la_nostra_societ%C3%A0_-180512655/?refresh_ce.

²⁰ U. RUFFOLO, *Intelligenza artificiale, Il diritto, i diritti, l'etica*, 2020, 10 s.

²¹ Undoubtedly, among the first to realise and perceive what would later be the developments between administrative activities and the use of technology, we remember, G. DUNI, voce Amministrazione digitale, in *Enc. Dir. Annali*, vol. I, 2007, Rome, 13 ss.; ID., *L'amministrazione digitale. Il diritto amministrativo nella evoluzione telematica*, 1992, Rome; again, on this point, please refer to A. MASUCCI, *Atto amministrativo informatico (voce)*, in *Enc. Dir.*, Agg.to, no. I, 2997, Milan, 221 ss. Even earlier, one cannot but recall M.S. Giannini himself, who in his well-known report on the state of the Italian administration stated that computerisation and the development of the administration were already closely linked. Refer to M. S. GIANNINI, *Rapporto sui principali problemi dell'Amministrazione sullo Stato*, in *Riv. Trim. dir. Pubbl.*, 1982, 722.

²² V. MANES, *L'oracolo algoritmico e la giustizia penale: al bivio tra la tecnologia e tecnocrazia*, in *Discrimen*, 15 May 2020.

²³ "In short, we can well say that what until a few years ago belonged to the world of science fiction is now a concrete fact. Every day, in fact, perhaps without even realising it, we interact with automatic machines that allow us to access the Internet to read the latest news, use e-mail, make reservations or purchases with a credit card, check our bank statement or the posts on our favourite social network, share a tweet, do an online search, make use of the services offered by Apple's Siri or Amazon's online shops, and use automatic dictation and simultaneous translation programmes". On this point, see A. TRAVERSI, *Artificial intelligence applied to justice: will there be a robot judge?*, in *Quest. Giust.*, 25 May 2021.

²⁴ In the United States, where the greatest diffusion and introduction of such tools was seen, the starting idea was to try to overcome and improve 'sentencing malpractice' through the introduction and aid of tools to support and assist the judge in prognostic evaluations, in order to move ever closer to transparent and more rational decisions. On the differences between the approach adopted in Europe and North America, see S. QUATTROCOLO, *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale "predittiva"*, in *Cass. Pen.*, 2019, 1748 ss.

no coincidence that there have been two recent regulatory interventions, which will be discussed below, that characterise the more cautious European attitude even more²⁵.

It is therefore the task and duty of legal scholars to confront the development and rapid evolution of these technologies²⁶, in order to understand the best possible use that can be made of them, the limits within which they can be applied, and above all, to identify in which phases of the criminal process and criminal justice - which involves various actors, such as the judge, the suspect, the defendant and other protagonists - they can be used²⁷.

The main area on which the thesis will focus will concern, in particular, the possibility of applying Artificial Intelligence tools in one of the most peculiar and delicate phases of the trial: in the sentencing phase. Furthermore, attention will be paid on risk assessment tools as artificial intelligence tools able to support the judge in assessing the dangerousness and capacity to commit a crime of an individual in the commensuration of punishment²⁸.

²⁵ In particular, reference is made to Regulation No. 679 of 27 April 2016 on the protection of natural persons (GDPR) and the European Ethics Charter for the use of A.I. On this point, for a more detailed analysis of the two legal acts, see Chapter V.

²⁶ 'Today we are facing the third revolution, that of artificial intelligence (AI). Activities requiring intelligence, hitherto carried out exclusively by people, can, to an increasing extent, be entrusted to machines, which have acquired the ability to reason, learn and act. Previously impossible applications - such as speech understanding, automatic translation, object recognition - are within the reach of every smartphone. An ever-widening set of functions can be delegated to intelligent technologies: automatic decisions, predictions about the behaviour of individuals and groups, control over workplaces and public spaces, biometric recognition, steering robots, driving autonomous vehicles, etc. This raises new legal problems, to which there are often no definitive answers. The practice of law is also susceptible to profound changes: expert systems, capable of applying formalised rules automatically, are flanked by machine learning functions, capable of extracting information from large masses of data and of building and applying predictive and decision-making models', see, G. SARTOR, *Intelligenza artificiale*, Turin, 2022.

²⁷ For an overview of the possible intertwining of A.I. and criminal law, see U. PAGALLO, *Saggio sui robot e il diritto penale*, in S. Vinciguerra-F. Dassano (eds), *Scritti in memoria di Giuliano Marini*, Naples, 2010, 595 ss.; J. CHARPENTIER, *Justice Machines. Racconto di fantascienza giudiziaria*, Macerata, 2015; S. RIONDATO, *Robot: talune implicazioni di diritto penale*, in P. MORO-SARRA, *Tecnodiritto. Temi e problemi di informatica e robotica giuridica*, Milan, 2017, 85 ss.; A. GARAPON - J. LASSÈGUE, *Justice digital. Revolution graphique et ropture anthropologique*, Paris, 2018; M. B. MAGRO, *Relazione su "Biorobotica, robotica e diritto penale"*, in *dirittopubblico.unipd.it*, 2018, 1 ss.; F. BASILE, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Diritto penale e uomo*, also published in *Diritto penale contemporaneo*, 29 September 2019, 1 ss. With specific regard to the relationship between criminal law and the phenomenon of automatic or semi-automatic cars, see A. CAPPELLINI, *Profili penalistici delle self-driving cars*, in *Diritto penale contemporaneo*, no. 2/2019, 326 ss. On the use of robots in medicine, see A. PERIN, *Standardizzazione, automazione e responsabilità medica. Dalle recenti riforme alla definizione di un modello d'imputazione solidaristico e liberale*, in *Rivista di BioDiritto*, no. 1/2019, 207 ss. On the distinct side of the use of AI as a tool for law enforcement, policing and predictive justice, v. M. LUCIANI, *La decisione giudiziaria robotica*, in *Rivista AIC*, no. 3, 2018, 872 ss.; C. BURCHARD, *L'intelligenza artificiale come fine del diritto penale? Sulla trasformazione algoritmica della società*, in *Riv. it. dir. proc. pen.*, 2019, 1909 ss.; A. GULLO, *Nuove frontiere tecnologiche e sistema penale: alcune note introduttive*, in *Riv. Trim - Dir. Pen. Cont.*, no. 2/2019, XI ss.; M.B. MAGRO, *Robot, cyborg e intelligenze artificiali*, in A. Cadoppi-S. Canestrari-A. Manna-M. Papa (eds), *Trattato di diritto penale - Cybercrime*, Turin, 2019; F. SGUBBI, *Il diritto penale totale*, Bononia, 2019, 40-44; V. MANES, *L'oracolo algoritmico e la giustizia penale*, 1 ss.

²⁸ Since the assessment of individual (social) dangerousness or of an individual's capacity to commit a crime is a type of assessment with which judges are confronted on a daily basis in the Italian system, it is considered to be useful also for the assessment of dangerousness in the pre-trial proceedings (Art. 274, lett. c c.p.) and when assessing the conditional suspension of the sentence (Art. 164 I paragraph c.p.) or the alternative measures to

Therefore, starting from questions which also include ethical and philosophical evaluations on the concept of dangerousness, an attempt will be made to trace a common thread among those instruments of this kind that already exist or are used, in order to understand and identify the crucial points of the decisional phase, where they could be applied. Undoubtedly, these new proposals go hand in hand with the need and demands emerging from the community, which is calling for greater efficiency and legal certainty, meeting new needs for speed and certainty in a judge's decision²⁹.

It is also necessary to take note that even within the Italian courtrooms, the idea of the introduction of such instruments is beginning to spread, which fascinate the scholar and at the same time show the weaknesses and difficulties that pertain to the tortuous path that the judge must take in making certain choices³⁰. The idea that artificial intelligence can shed a new light on the value that lawyers, magistrates, court clerks and practitioners in general bring to the functioning of the justice system is therefore beginning to spread more and more.

In the face of the silent change of a justice system that is becoming more technical³¹, we watch as mere spectators to changes that are only partly anticipated by non-European panoramas. The idea of a judgement becoming *more exact* passes through the meshes of a difficult balancing act between opposing values and guarantees, the centre of a new battlefield of debate involving legal interpreters, philosophers and legal computer scientists³².

In the face of the *silent change* of a justice system which is becoming *more technical*, we stand as mere spectators to changes that are only partly anticipated by non-European vistas. The vision of a judgement becoming *more exact* passes through the meshes of a difficult balancing act between opposing values and guarantees, the centre of a new battlefield of debate involving legal interpreters, philosophers, and legal computer scientists.

In conclusion, the precise choice of focusing the analysis of this thesis on the sentencing phase is developed following two directives: firstly, investigating the concept and assessment

detention, or, lastly, when deciding on the quantity of the punitive treatment under Art. 133 c.p., paragraph II of the Criminal Code on the capacity to commit offences.

²⁹ D. POLIDORO, *Tecnologie informatiche e procedimento penale: la giustizia penale "messa alla prova dall'intelligenza artificiale*, in *Archivio penale*, No. 3, 2020, 13 s.

³⁰ In fact, in a decision taken by the Court of Milan, reference was made to the use of a tool for the assessment of recidivism risk for the application of the security measure. In the specific case it concerned only a particular type of crime, such as sex offenders. Court of Milan, 19 April 2016. These elements are shared by the man during individual interviews or during the recidivism risk survey, carried out with North American instruments, which indicate an overall 'medium-high' level (about 25%) of possibility of committing a new sexual offence in the five years after detention.

³¹ C. COSTANZI, *Big data e garantismo digitale. Le nuove frontiere della giustizia penale nel XXI secolo*, in *Leg. Pen.*, 2019, 3 ss.

³² C. CASTELLI–D. PIANA, *Giusto processo e intelligenza artificiale*, *Santarcangelo di Romagna*, 2019, 10.

of an individual's dangerousness and, secondly, on how the A.I. tool can be useful in the phase of the commensuration of the sentence for the purposes of choosing the best sanctioning treatment.

The interest in the subject is derived from the particular attention in relation to the human judgement, with the skills and limitations of the same, which may turn towards something that detaches itself from the canons of statistics and mathematics to represent a convoy of a series of elements that lead to the final decision. And it is precisely this aspect which attracts the use of certain tools that would seem to give the jurist a more solid basis through which to imagine being able to 'improve' human judgement, making it, precisely, less fallible.

Therefore, starting from these premises, an attempt will be made in this first part of the paper to follow the thread that will guide the reader along the path, initially merely descriptive, and then to focus attention on the possible applications that we propose to describe. In the continuation, an attempt will be made to traverse and dissect transversally the enormous breadth of the theme, trying to focus more on unravelling the most relevant knots that arise when these instruments meet with criminal justice. Indeed, by abandoning the purely unidirectional study of criminal justice issues, it is necessary to adopt a transversal approach that analyses the topic while maintaining a multidisciplinary approach that intersects the study of the main issues of other disciplines such as legal informatics, ethics and law. In the face of great questions about 'who decides' and 'how decisions should be made', the entry or approach of these new tools counterbalances a need for improvement and 'speed' in decision-making.

In the hard research of a point of equilibrium between the fascination that these tools and the impact on criminal justice, it assumes as the bearers of a 'new industrial revolution' capable of raising and improving the levels of efficiency, savings and safety of society and the habits of life, restoring a better existence to mankind, an enormous and debated social conflict arises, which gives rise to questions and doubts of an ethical and social nature that are still unresolved. Within this copious and difficult debate, one cannot disregard the increasingly invasive role of technology and technocrats who strongly support the entry of these tools even into the courtrooms. The questions and doubts undoubtedly possess a *disquieting nature* that places the jurist in the necessary condition to unravel the issues and place himself in front of the risks in

order to face them as best he can, trying to understand their possible usefulness and the appropriate regulation³³ and 'governance' over them³⁴.

In conclusion, taking as our starting point precisely the oldest and most difficult of questions in the field of robotics³⁵, "*Can machines think?*", we consider how, in part, machines are no longer merely tools used by man to construct concrete objects and transform them, readjusting them to his needs, to solve problems, but require a conceptual leap that may apparently be much more challenging. However, responding to the need to overcome such questions in order not to infiltrate fields that call philosophy, gnoseology (and other sciences) into play, one must surely imagine that since there is no circumscribed and objective definition of human thought, it is also necessary to consider the very limitations of human language: indeed, the natural language with which one expresses oneself is not always suitable to fully represent the functions performed by a complex organism such as a human being, but an ineradicable grey area of ambiguity will still remain that will not allow the nuances of thought³⁶ to be fully and completely captured. Here, an attempt will be made to understand whether certain tools can actually support the human being in certain peculiar evaluations that imply, in themselves, thought and reasoning, and to assess the advantages and risks of approaching such tools in the delicate decision-making phase³⁷.

Lastly, in an attempt to answer the research questions posed at the beginning of the paper, it is anticipated as of now that not all of them have been answered in a definite or unchangeable manner; what is more, however, is to be able to provide a complete overview of the state of the art of these instruments; Indeed, the results produced by the use of such instruments in the US

³³ "Is it conceivable to appeal, again, to the traditional hard-law regulatory apparatus, which, by its very nature, favours a 'reactive' perspective, i.e. one capable of acting only on the undesirable consequences arising from AI penetration? Or is it necessary to move in the direction of a 'proactive' approach, capable, that is, of intercepting risks and governing the problems in advance?", thus, on the point, C. PIERGALLINI, *Intelligenza artificiale*, 1746 who questions the possibility and modalities of regulation. Moreover, on this subject, G. MOBILIO, *L'intelligenza artificiale e i rischi di una "disruption" della regolamentazione giuridica*, in *Rivista di BioDiritto*, No. 2, 2020, 401 ss.

³⁴In doctrine, there is a call for European intervention for an effective regulation of artificial intelligence, which is gradually taking place through the introduction of soft law instruments. On this point, see G. PASCERI, *Intelligenza artificiale. Algoritmo e machine learning*, Milan, 2021, 53 ss.

³⁵ A. TURING, *Mechanical Intelligence*, North Holland, 1992.

³⁶ On the further implications and problems that arise when approaching human and machine thinking, for an interesting introductory section see G. D'ACQUISTO, *Intelligenza artificiale. Elemento*, Turin, 2 ss. In particular, the author believes that 'for almost a century now, mankind has intuited this great new field of application for machines, and today in particular feels its urgency given the abundance of information in which he lives and the ease with which it can be generated and exchanged on a global scale. Information, we could say with all the ambiguity of natural language, feeds our thoughts, and to be able to generate it in an automated form would constitute for mankind a leap forward in knowledge of the world never before experienced. A new *renaissance*, or *renAIssance*, as we call it, right with the capital A and I of Artificial Intelligence'.

³⁷ See D. DE KERCKHOVE, *Algoritmo, Big data e il sistema legale*, 53.

criminal justice system undoubtedly provide the first answers to these questions, which will have to be taken into account when in Italy, but more generally in Europe, we are faced with instruments or 'predictive systems' that are as innovative as they are, at the same time, dangerously in conflict with the constitutional rights placed to protect the accused in the criminal justice system.

1.1 *The boundaries of the study: the continuation of the premise*

On closer inspection, the study on these issues will try to start from the *fascination* they hold and the state of the art that has brought these A.I. tools to the centre of the debate. However, if one juxtaposes Artificial Intelligence and criminal law today, one can see how doctrinal reflections propagate in different directions. The focus of today's study, for systematic reasons but also for the sake of in-depth analysis, will be on understanding how predictive analysis tools, in particular algorithms, can act as a support for the judiciary.

Indeed, being able to consider and analyse all the tools or individual models of machine learning is a very difficult undertaking that would require a single study on each of them. Therefore, the analysis will be limited to a theoretical descriptive investigation, attempting to provide the legal parameters within which one can 'start' to hypothesise the application of such tools, in order to be able to understand the possibility of introducing such tools and, at the same time, the various reasons and limitations.

Indeed, the reflection will only move from a legal and not a technical perspective.

On closer inspection, an attempt will be made to address the issue through multifaceted research, which sees the intersection of different fields such as criminal law, criminology, but also the philosophy of law, which is responsible for addressing the ethical issues related to the concept of decision-making. In the criminal justice system, the analysis of dangerousness and how the judge should question himself in order to give an answer on sanctioning treatment with a view to the future, has always been at the centre of the debate. For this motive, one of the goals has always been creating quantitative models to predict risk (of recidivism and violent behaviour): today, such models can benefit on the most recent achievements in computer science³⁸.

In fact, in a highly discretionary area of human activity, such as criminal justice, artificial intelligence may appear a potential, effective, solution to many recurrent conundrums. Moving from the current achievements of artificial intelligence, the thesis offers a reflection on the

³⁸ S. QUATTROCCOLO, *Sui rapporti tra pena, prevenzione del reato e prova nell'era dei modelli computazionali psico-criminologici*, in *Teoria e critica della regolazione sociale*, Vol. 1, No. 22, 2021.

possible role of these technologies, in order to understand how they can (maybe) contribute to restore some effectiveness to a system that has lost (from some parts) credibility and trustworthiness.

The idea of being able to make a 'right' decision or an 'exact' prognostic assessment of criminal dangerousness would seem to represent not only the wish of those involved in criminal proceedings as defendants, but also the ambitious goal of the human being who is striving more and more to overcome boundaries that are being pushed further and further out³⁹. It is therefore necessary, even in the field of criminal science, to take note of the fact that the until a few years ago even unimaginable progress that computer science has achieved in various fields is increasingly manifesting itself, even in the field of law, thanks to the development of software that is able to 'produce algorithmic decisions'.

In particular, criminal law must try to keep pace with the unstoppable progress of technological evolution in order to address the new legal issues and problems that arise⁴⁰.

In an attempt, therefore, to read the subject through a twofold key, an attempt will be made to look at algorithms as tools that could perhaps confer greater certainty or 'exactitude' to the law and guarantee, at the same time, greater predictability in the decision; at the same time, an attempt will be made not to lose sight of the guarantees in an attempt to envisage a way to, firstly, identify the major points of friction and, secondly, to try to find a way to overcome them. Moreover, in an initial attempt to delimit the field of investigation of this research and elaboration, we started from a premise: from the choice of incrimination, to the formulation of the precept, from the construction of offences of danger to the ascertainment of the causal link, from recidivism to punishment and security measures, various types of prognosis run through the entire penal system. Indeed, prognostic judgements are at the heart of the penal system.

Certainly what is important is that in a punitive system inspired by the retributive ideal, prognoses represent almost a 'foreign body': indeed, if one were content to establish, through the codified parameters, what is the most appropriate and proportionate measure to be inflicted on the offender in relation to the disvalue of the fact committed, one would probably end up losing sight of a type of assessment that, on the contrary, is entirely central and cannot be avoided.

³⁹ See C. ZINGALES, *Risk assessment tools: una nuova sfida per la giustizia penale?*, in *DPC*, 12th December 2021.

⁴⁰ In an attempt to address problems 'similar to those that have characterised other technological 'transitions': verifying the suitability of existing rules to apply to new technologies, so as to assess whether it is appropriate for legislators to coin ad hoc, new rules, or to persist, not without possible forcing endorsed, perhaps, by case law, in the application of pre-existing rules', M. BASSINI - L. LIGUORI - O. POLLICINO, *Sistemi di Intelligenza Artificiale: responsabilità e accountability. Verso nuovi paradigmi?*, in F. Pizzetti (ed), *Intelligenza artificiale*, 334.

In this research work, therefore, an attempt will be made to take into consideration those algorithms that operate in an attempt to provide a general datum (a score) on the future commission of an offence, following the ascertainment of responsibility for a criminal offence that has been committed and already ascertained. Undoubtedly, this question is of great and considerable interest to the criminal lawyer, since the risk of future offences being committed is one of the elements on the basis of which to define an appropriate response to the offence, in the sense of a penalty treatment that can be defined as truly individualised⁴¹.

2 *Artificial intelligence: a brief outline of its origins and evolution*

In attempting to provide a definition of Artificial Intelligence⁴², it should be pointed out that there is no uniform notion of the concept, as it has many facets and encompasses a huge number of meanings⁴³. It can certainly be premised that the term was coined by John McCarthy⁴⁴ in

⁴¹ The individualisation of criminal penalties has always been at the centre of scientific debate, now constituting a fundamental principle of criminal law. Although it is not expressly included in the Constitutional Charter, it has become part of it following numerous judgments of the Constitutional Court that, reasoning on the scope of Articles 25 and 27 of the Constitution, have affirmed that the legality of punishment cannot disregard its individualisation. Individualisation, in fact, "stands as the natural implementation and development of constitutional principles, both of a general order (principle of equality) and pertaining directly to criminal matters": Constitutional Court, Sentence No. 50 of 2 April 1980. Recently, the C. cost. has expanded on the principle of individualisation of punishment, especially with reference to the granting of prison benefits (cf. C. cost., sent. 15 February 2022 no. 33). This principle has also been used as a tool to unhinge the remaining automatic sanctions within the system (see, most recently, Constitutional Court, judgment no. 56 of 31 March 2021). On this subject see the reflections of A. CARCANO, *Automatismi: tra ragionevolezza e individualizzazione della pena*, in *Forum di Quaderni costituzionali – Rassegna*, 2021, 4. The principle of individualisation of punishment has, moreover, been the subject of a very recent and rich study, in doctrine, by M. VENTUROLI, *Modelli di individualizzazione della pena. L'esperienza italiana e francese nella cornice Europea*, Turin, 2020.

⁴² It seems necessary to recall how the founding father of Artificial Intelligence can be traced back to Alan Turing and, in particular, the moment of its birth in 1950 with the publication in *Mind* magazine of an article by him, entitled "*Computing machinery and intelligence*". In particular, the famous incipit of that article already mentions issues that have been addressed by research in the years to follow. It thus reports "*I propose to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous, if the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, 'Can machines think?' is to be sought in a statistical survey such as a Gallupo poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words. [...] May not machines carry out something which ought to be described as thinking but which is very different from what a man does?'", as also reported G. D'ACQUISTO, *Intelligenza artificiale. Elementi*, 2.*

⁴³ One of the most accepted definitions today is the recent one as defined by the Council of Europe. Specifically, it is a "*set of sciences, theories and techniques, the purpose of which is to reproduce, through the machine, the cognitive capabilities of a human being*". Thus, European Commission for the Efficiency of Justice (CEPEJ), *Ethical Charter for the Use of Artificial Intelligence in Justice Systems and their Environment*, App. III, Glossary, 47.

⁴⁴ "[AI] is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable", J. McCarthy, *What Is Artificial Intelligence*, rapp. tecn., Stanford University, 2007.

1955⁴⁵, but the expression “Artificial Intelligence”⁴⁶ was used, for the first time, by Marvin Minsky⁴⁷ in a seminar at Dartmouth⁴⁸. The history and origins of Artificial Intelligence⁴⁹ lead to a fascinating landscape that is today at the centre of the debate on technological development and progress involving a wide variety of actors, such as companies, businesses, and government management⁵⁰.

On closer inspection, Artificial Intelligence is a specific branch of legal informatics that deals with the design, creation and programming of systems capable of creating machines that try to come ever closer to human behaviour⁵¹. It should be noted, however, that Artificial

⁴⁵ At a seminar he organised at Dartmouth College in Hanover, New Hampshire, USA, the term was coined that gave that branch of research its true autonomy from that moment on. Specifically, it was defined as 'a process of allowing a piece of equipment to behave in ways that would be called intelligent if it were a human being behaving in the same way', thus J. MC CARTHY-M. MINSKY-N. ROCHESTER-C. SHANON, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, Stanford, 1995.

⁴⁶ Although there is no specific definition of Artificial Intelligence⁴⁶, either at a legislative or jurisprudential level, most scholars accept the broadest definition given to the concept. In fact, according to an internationally accepted definition, "Artificial Intelligence" is that discipline, “*belonging to computer science, which studies the theoretical foundations, the methodologies and the techniques which allow the design of hardware systems and software programme systems providing the computer with performances which, to a common observer, would seem to be of exclusive pertinence to human intelligence*”. In realtà, occorre notare come manchi una definizione univoca e condivisa già del concetto di “intelligenza”. One of the most authoritative introductory texts on the subject, the Oxford Companion to the Mind, opens its discussion of the entry 'intelligence' by saying that 'there are innumerable tests available to measure intelligence, but no one knows for sure what intelligence is, and even no one knows for sure what the available tests measure', thus on the point G. SARTOR, *Intelligenza artificiale e diritto*, 2022, 1. See also R. L. GREGORY, «Intelligence», in R. L. Gregory (ed) *The Oxford Companion to the Mind*, Oxford, 1987, 375-379.

⁴⁷ For a careful reading on the subject, please refer to a paper by M. MINSKY, *Steps toward Artificial Intelligence*, in *Proceeding of the IRE*, 10 January, 1979. The definition he gave to Artificial Intelligence was the 'science that makes machines do things that would require intelligence if they were done by humans'.

⁴⁸ Actually, the origins of the first debate on this topic were even older; in fact, scientific and technological research on AI began in the 1940s and 1950s. As early as 1943, Walter Pitts and Warren Sturgis McCulloch (two collaborators of Norbert Wiener, the inventor of cybernetics) showed how networks of artificial neurons could process information, starting the research on neural networks. Thus, G. SARTOR, *L'intelligenza artificiale e il diritto*, c25.

⁴⁹ "There is no complete consensus on what the term 'artificial intelligence' means, but it could be argued that the term describes the possibility that machines, to a certain extent, 'think', or rather mimic the human thinking, based on learning and the use of generalisations, that people use to make everyday decisions", on the point thus, J. N. FENOLL, *Intelligenza artificiale e processo*, Turin, 2018, 8 ss; On this point, it is also useful to read R. LOPEZ DE MANTARAS BADIA-P. MESEGUER COLZALEZ, *Intelligenza artificiale*, Madrid, 2017, 18 ss.; and again, J. KAPLAN, *L'intelligenza artificiale. Guida al futuro prossimo*, Rome, 2017, 15 ss.

⁵⁰ Interestingly, however, an attempt has been made to provide a definition also at supranational level. European Commission COM (2020) 64, Final, White Paper on Artificial Intelligence - A European approach to excellence and trust, 2020. recently one of the White Papers on Artificial Intelligence, published by the European Commission in 2020, firstly refers to Somalvico's definition and secondly pragmatically tries to point out the problems associated with the cognitive development of Artificial Intelligence.

⁵¹ A more specific definition was provided by CEPEJ, which outlined it as 'Artificial intelligence is defined as a set of scientific methods, theories and techniques, the goal of which is to reproduce through a machine the cognitive abilities of human beings. Current developments seek to make machines capable of performing complex tasks typically performed by humans. The term is criticised by experts who tend to distinguish between strong (or maximum) artificial intelligence, i.e. when it is capable of performing complex tasks in a fully automated manner, and weak or moderate artificial intelligence, i.e. when artificial intelligence requires learning. CEPEJ, Ethical charter on the use of artificial intelligence in judicial systems and their environment, 2018.

Intelligence is recognised as an autonomous branch⁵² which has links with other subjects such as mathematics, computer science, cognitive science, neurobiology and philosophy⁵³.

More specifically, the term Artificial Intelligence indicates '*those technological systems that exhibit intelligent behaviour by analysing their environment and performing actions, with a certain degree of autonomy, to achieve specific goals*'⁵⁴.

As already mentioned, it has always been quite difficult to formulate an unambiguous definition⁵⁵ of Artificial Intelligence⁵⁶; in many cases, the most complex question has concerned the definition⁵⁷ and limits on the concept of 'intelligent'⁵⁸. In Italy, one of the

⁵² Actually, on closer inspection, Artificial Intelligence is the subject of several disciplines, including philosophy, mathematics, medicine, psychology and linguistics. In particular, AI has drawn inspiration from all the research just mentioned but has added an engineering aspect to these: AI does not only want to study intelligence, but aims to build it, to bring intelligent artefacts to life. The engineering objective of AI does not exclude that it can contribute to the knowledge of human intelligence', G. SARTOR, *Intelligenza artificiale e diritti*, 2022, 2.

⁵³ F. RIGUZZI, *Introduzione all'Intelligenza artificiale*, 11 May 2021.

⁵⁴ Communication from the European Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Artificial Intelligence for Europe, 25 April 2018, available at [webhttps://ec.europa.eu/transparency/regdoc/rep/1/2018/IT/COM-2018-237-F1-IT-MAIN-PART-1.PDF](https://ec.europa.eu/transparency/regdoc/rep/1/2018/IT/COM-2018-237-F1-IT-MAIN-PART-1.PDF).

⁵⁵ Today, another definition provided by the European Ethical Charter on the Use of Artificial Intelligence in Justice Systems and Related Fields is accepted; according to this source, artificial intelligence consists of a "set of scientific methods, theories and techniques aimed at reproducing by means of machines the cognitive capacities of human beings", as stated in Appendix III of the European Ethical Charter for the Use of Artificial Intelligence in Justice Systems and Related Fields, adopted by the Commission on the Efficiency of Justice (CEPEJ) in 2018. Please refer to Chapter V for a more in-depth analysis.

⁵⁶ It has also been observed that 'as suggestive as it may be to speak of Artificial Intelligence, it should be noted that, in reality, "Little, apart from speculation and naive thinking, links today's work in the field of A.I. to the mysterious mechanisms of the human mind; in reality, at least at this stage, it is an engineering discipline with more of a metaphorical and 'inspirational' relationship with biological organisms", all the more so since intelligence (that of human beings, even before that of machines), although the subject of numerous studies by psychologists, biologists and neuroscientists, still constitutes an indeterminate concept. For this and other reasons, A.I. researchers sometimes prefer to speak - rather than of intelligence - of rationality, where 'rationality' means the capacity to choose the best action to take in order to achieve a given objective in the light of certain criteria for optimising the available resources', see F. BASILE, *Intelligenza Artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Diritto penale uomo*, which in turn refers to and quotes, S. RUSSEL-P. NORVIG, *Artificial Intelligence: A modern approach*, prentice Hall, III ed., 2009, 36 ss.

⁵⁷ They note the absence of a definition, among many, M. B. MAGRO, *Biorobotica, robotica e diritto penale*, in D. Provolo - S. Riondato - F. Yenisey (eds), *Genetics, Robotics, Law, Punishment*, Padua, 2014, 510 s.; R. CALO, Artificial Intelligence Policy: a Primer and Roadmap, in *University of Bologna Law Review*, 3:2, 2018, p. 184; C. TREVISI, *La regolamentazione in materia di Intelligenza artificiale, robot, automazione: a che punto siamo*, in *Medialaws*, 21th May 2018, 1.

⁵⁸ Some doubts and questions can be found among scholars. In particular, on the very concept and definition of 'intelligent', the Oxford Companion to the Mind, in defining 'intelligence' stated that: "countless tests are available to measure intelligence but no one knows for sure what intelligence is, and even no one knows for sure what the available tests measure". On this point, see L. GREGORY, *Innumerable tests are available for measuring intelligence, yet no one is quite sure of what intelligence is, or even of just what is that the available tests are measuring*, voce "Intelligence", in *The Oxford Companion to the Mind*, Oxford, 1987, 375. On the same concept, the definition given by an American psychologist appears interesting. Referring to a study on the concept of 'intelligent', she concluded that 'intelligence is not only the ability to reason, plan, solve problems, think abstractly, converse and understand complex ideas and learn from experience, but it is characterised by abilities that constitute adaptability, wit, intuitive and profound capacity to understand things and events that we perceive, being able to instinctively attribute meaning to them and being able to behave accordingly in a more or less shrewd manner', refer a L.S. GOTTFEDSON, *Mainstream Science on Intelligence*, in *Wall Street Journal*, New York, 13 December 1994.

pioneers of Artificial Intelligence has defined it as "*that discipline, belonging to computer science, which studies the theoretical foundations, methodologies and techniques that enable the design of hardware systems and systems of software programmes capable of providing the computer with performances that, to a common observer, would appear to be the exclusive domain of human intelligence*"⁵⁹.

The idea behind the design of such 'intelligent' machines is to create systems capable of solving complex problems. In this regard, there are various artificial intelligence systems that can be diversified according to the different operating technologies or in relation to their field of use or their incorporation into different hardware devices. On closer inspection, A.I. models may consist of software acting exclusively in the virtual world (e.g. facial and voice recognition systems), or they may be embedded in hardware devices and act synergistically with them⁶⁰.

The fields of application of Artificial Intelligence today are manifold and are having a great impact on people's lives and on the use of such systems also in public administrations, but also in certain aspects of justice⁶¹.

As there is no universal definition of the term Artificial Intelligence today, in an attempt to provide a more complete picture and a definition of the very extensive field in which it extends, scholars have tried to provide and identify the main characteristics of the different tools of when it comes into contact with artificial Intelligence⁶².

Undoubtedly, a peculiar aspect with which the jurist is confronted and clashes nowadays, is the confrontation and relationship with an 'unknown entity', whose first obstacle is undoubtedly represented by the terminology that is used, which, in this head, is rather 'opaque'⁶³ and in some cases indecipherable.

Artificial intelligence, in its general characteristics, presents certain features that then allow distinctions to be made between strong and weak artificial intelligence systems⁶⁴. To date, the

⁵⁹ M. SOMALVICO, *Intelligenza artificiale*, in *Scienza&Vita*, no. 8, 1987.

⁶⁰ See, V. GUARRIELLO, *L'intelligenza artificiale tra profili giuridici ed alcune delle più attuali applicazioni al servizio della società*, in *ARSG*, 19 November, 2021, s.

⁶¹ In Italy, for example, some scholars at the University of Florence are currently working on an experimental basis on the study of some algorithms aimed at identifying a probability value with which a dispute could, albeit in principle, be settled out of court. C. CASTELLI-D. PIANA, *Giustizia predittiva*, 11.

⁶² D. POLIDORO, *Tecnologie informatiche e procedimento penale*, 3.

⁶³ E. CALZOLAIO, *Intelligenza artificiale ed autonomia della decisione: problemi e sfide*, Milan, 2020, 1. On this point, see also I. GIUFFRIDA - F. LEDERER-N. VERMEYS, *A legal perspective on the trials and tribulations of A.I.: how artificial intelligence, the Internet of Things, Smart Contracts, and other technologies will affect the Law*, in *Case Western Reserve Law Review*, 2019, 747 ss.

⁶⁴ According to the characterisation given by John Searle, a distinguished scholar of the language of the mind, strong AI is based on the assumption that computers are also capable of cognitive states and thought (in the way a human being is endowed with them) and consequently proposes to build artificial minds. For strong AI, 'the appropriately programmed computer is really a mind, i.e. it can literally be said that computers equipped with the right programmes understand and have cognitive states'. Weak AI, on the other hand, proposes to build artificial

most widespread A.I. systems are those that present characteristics that fall more within the weak artificial intelligence systems. Clearly, the idea of trying to create machines capable of thinking and able to bring about certain results, trying to bring the machine as close as possible to human behaviour and the human mind, is one of the objectives that remains central to research today⁶⁵.

In conclusion, an initial difficulty encountered by jurists, but above all by A.I. scholars, is immediately apparent: indeed, the difficulty of being able to provide a definition of the concept⁶⁶ is one of the most difficult problems for today's jurist to deal with the themes and issues arising from the first applications of these instruments. Indeed, the jurist must be able to deal with concepts that present a sufficient degree of precision to enable the addressees of the rules or those who must ensure their implementation to distinguish the objects or phenomena to which those concepts apply⁶⁷.

2.1 *Machine learning, deep learning and dynamic risk analysis systems*

On closer inspection, it is important to introduce the topic of machine learning and deep learning, which represent 'the most advanced tip of artificial intelligence'⁶⁸. Machine learning is a peculiar branch of computer science that evolved from the study of pattern recognition and

systems capable of performing complex tasks, systems that can mimic (simulate) aspects of human cognitive processes, but cannot reproduce those same processes. See J. R. SEARLE, *Minds, Brains and Programs*, in D. C. Dennet (ed), *The Behavioural and Brain Science*, 1980), 417-57. Other authors (futurologists or science fiction writers) speak of strong AI in a meaning approaching that of 'artificial general intelligence', to refer to the goal of realising artificial systems whose cognitive skills are general, and potentially reach or exceed human capabilities.”⁶⁵ Since ancient times, human beings have always been fascinated by the idea of creating, albeit artificially, human entities capable of thinking. Indeed, on this very point, 'Intelligent automata can already be found in the myths of ancient Greece: Pygmalion sculpted Galatea, a living statue (thanks to divine intervention); the god Hephaestus could create animate bronze beings, like Talos, the legendary guardian of Crete. Moving from myth to mechanical engineering, we can mention in antiquity the automata built by Heron of Alexandria (who lived in the first century and invented, among other things, the steam engine), used to animate deities in temples. In more recent times, we can recall the myth of the Golem of Prague, created to defend the Jewish ghetto from anti-Semitic attacks, which escaped the control of its creator', so reports G. SARTOR, *L'intelligenza artificiale e il diritto*, 23.

⁶⁶ A different definition was provided by the High Level Expert Group on AI (AIHLEG, set up by the European Commission) in its report on the elaboration of a European strategy on AI, which preceded the Draft Regulation; thus, AI-HLEG, High-Level Expert Group on Artificial Intelligence, A definition of AI: Main capabilities and scientific disciplines, European Commission, 2019. The original version: “*Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions*”.

⁶⁷ Cfr. “Se non abbiamo un concetto condiviso di intelligenza, o comunque non è possibile stabilire in modo preciso che cosa sia intelligente e che cosa non lo sia, come possiamo distinguere i sistemi informatici “intelligenti” da quelli privi di intelligenza, al fine di applicare solo ai primi le norme sull’IA?”, così le preoccupazioni mosse da G. SARTOR, *L'intelligenza artificiale e il diritto*, 7.

⁶⁸ G. D'ACQUISTO, *Intelligenza artificiale*, op. cit., 198.

computational learning theory in the field of artificial intelligence. Machine learning systems are present in many aspects of our daily lives⁶⁹. Indeed, they are used to a great extent and can be found in many web platforms. They are mostly systems that are able to 'learn' and 'learn' from the experience and reality they come into contact with'. Machine learning systems refer to systems that 'improves its performance of future tasks after making observations'⁷⁰. More specifically, machine learning consists of an algorithm that is able to learn from the data within it; it is then able to 'refine' its knowledge when it 'acts and interacts' with experience. Learning, in particular, is then aimed at prediction, at solving cases other than those analysed that may occur in the future. Moreover, this process performed by the machine is also referred to as 'generalisation': given a set of initial information, a rule is then extrapolated that is suitable for predicting and solving future cases that have not yet been analysed.

The most widely used and popular definition of machine learning today was provided by Tom Mitchell⁷¹ who defined it as "*A programme is said to learn from experience E with reference to some class of task T and with performance measurement P, if its performance in task T, as measured by P, improves with experience E*".

As the same of the algorithms, machine learning⁷², through the processing of data within it, is able to 'learn' through impact with experience. It is believed that machine learning tools are useful precisely because they 'learn' from the past' through prediction, thus improving subsequent cases that occur in the future. The process carried out by machine learning systems is also referred to as the 'generalisation process' in which it is based on a set of starting information, from which a 'general' rule is then extrapolated, which is then able to predict and

⁶⁹ On these self learning machines, see P. NORVIG, *Macchine che apprendono*, in D. Heaven (ed), *Macchine che pensano*, 31 ss.; for an analysis projected onto the terrain of legal consequences, v. S. BECK, *Intelligent agents and criminal law - Negligence, diffusion of liability and electronic personhood*, in *Robotics and Autonomous Systems*, 2016, 138 ss.

⁷⁰ J. RUSSEL-P. NIRVIG, *Artificial Intelligence. A modern approach*, 3° ed., Prentice Hall, Englewood Cliffs, 2010, 693.

⁷¹ It should be remembered that the term 'machine learning' was actually first coined by Arthur Lee Samuel in 1959. For a more complete analysis in one of his early writings on the subject, see a A. L. SAMUEL, *Some Studies in Machine Learning Using the Game of Checkers*, in *IBM Journal of Research and Development*, No. 44, Issue:1.2.

⁷² To be adjusted and put in the bibliography: Machine learning algorithms stand out for their ability to 'learn' the programme automatically from observing data. In other words, although they do not replicate the features of the human cognitive system, they are nevertheless able to implement their own performance with regard to future tasks, after having observed reality. Those who programme it are unable to predict changes in a given situation over time: in other words, one can 'see' the result offered by the system, but one cannot understand how that result was achieved. On these aspects, see J. COPELAND, *Artificial Intelligence*, in S. GUTTENPLAN, *A companion to the Philosophy of Mind*, 1996, 124. It should also be noted that, not infrequently, especially in the area of predictive functions, algorithms can hand down decisions steeped in 'bias': these are 'dysfunctionalities' that algorithms learn, mostly, from those who programme them and from the data from which they 'learn'. As noted, 'algorithms have learned to be biased by us'. On this point, A. VESPIGNANI, *L'algoritmo e l'oracolo*, 103.

solve future cases and for which the rule is therefore applicable⁷³. These systems are taken into consideration by the paper because it is precisely such systems, equipped with an artificial neural network, that are able to learn and store an unlimited amount of data and then process it to produce a result that depends on the input and the question to be solved⁷⁴. Basically, machine learning systems analyse and explore data.

More specifically, they work on data and process it by means of an already labelled 'training set' or process it in data autonomously in order to derive patterns and predictive models. The operation of such systems works through an inductive logic: the machine learning algorithm, in fact, first analyses and 'observes' a set of data and elaborates certain rules that it standardises⁷⁵. Subsequently, by observing other data, it is able to recognise connections and modify its knowledge⁷⁶.

Machine learning systems are generally based on different types of methods: inducing decision trees, e.g. statistical methods (e.g. regression), association rule learning, artificial neural networks, bayesian networks, genetic algorithms, support vector machines, etc. More generally, it is possible to identify three main types of machine learning: supervised learning, in which the A.I. system observes a few examples of unput-output pairs and learns the function that maps from input to *output*; reinforcement learning, which differs from standard supervised learning in that correct *input-output* pairs are never presented; and, lastly, unsupervised

⁷³ A. M. MAUGERI, *L'uso di algoritmi predittivi per accertare la pericolosità sociale: una sfida tra evidence-based practices e tutela dei diritti fondamentali*, in *Arch. Pen.*, No. 1, 2021.

⁷⁴ On the subject of instruments used in the formulation of judgments, see A. TRAVERSI, 5. The author examines these instruments by evaluating them as possible applications in the formulation of certain types of judgments, such as in civil cases concerning compensation for damages resulting from road accidents and also in the taxation sector, in assessing how some of them can automatically resolve certain disputes. The author undoubtedly considers the major problematic profiles that arise with regard to the possible application of such tools in the criminal justice sector for three orders of reasons: "First, because the most frequently used means of proof in criminal proceedings for the ascertainment of material facts is testimony, and a computer would encounter serious difficulties in judging whether a witness has told the truth, been reticent or lied. Secondly, because the criteria for evaluating evidence are multiple and not predetermined, so that, especially in a circumstantial trial, it would be even more difficult for a computer to establish whether certain clues are to be considered 'serious', 'precise' and 'concordant', so that the existence of a fact can be deduced from them, as prescribed by Art. 192(2) of the Code of Civil Procedure. Lastly, since the computer is programmed to provide certain answers, it cannot have any doubts, whereas in our legal system there is the principle, enshrined in Article 533, paragraph 1 of the Code of Civil Procedure, according to which the judge pronounces a sentence of conviction "if the defendant is guilty of the offence charged against him beyond any reasonable doubt" and must instead acquit if that parameter is not exceeded".

⁷⁵ For a more specific definition of an algorithm, see M. SIPSER, *Introduzione alla teoria della computazione*, ed. it., C. DE FELICE - L. GARGANO -P. D'ARCO (eds), Santarcangelo di Romagna, 2013, 191 ss. In particular, an algorithm is defined as a "finite sequence of repeatable and unambiguous instructions, indicating a combination of actions to be performed to solve a problem". It has the following characteristics: finiteness, in the sense of leading to the solution through a finite number of sequences; generality, since it must be able to solve a class of problems; univocity, since the operations must always be carried out in the same way; repeatability, whereby given the same inputs the algorithm must provide the same outputs.

⁷⁶ R. CERVELLI, *Machine learning: cos'è e come funziona l'apprendimento automatico*, 9 May 2019.

learning, which is a particular type of machine learning where data sets with no specified structure are used, and no explicit feedback is provided to the system.

The machine can work with and processing a large set of data and, consequently, the greater the number of data and sources, the more specific will be the ability of the algorithm to lead to exact predictions. The element that qualifies machine learning systems as 'intelligent' agents is the discovery of an underlying law; in particular, the agent hypothesises the existence of a law that produced those observations and attempts, on the basis of those observations, to discover it and use it for future explorations of the world⁷⁷.

Undoubtedly, a peculiar aspect, which we merely mention here, is strictly related to the risks and obscure points of such systems; in machine learning systems, a phenomenon is considered to be a black box in which the agent ignores the internal mechanisms, being able and limited only to observing the values of certain quantities that he considers to be input variables, and those of other quantities that he considers instead to be 'output values'. It therefore remains the agent's task to find the specific function that links those inputs and outputs. In particular, we are dealing with simple learning systems (paradoxically simpler than human abstraction) which, however, have the advantage of being able to be applied indiscriminately to any phenomenon, from the prediction of stock market prices to the analysis of a meteorological phenomenon⁷⁸.

⁷⁷ "At first glance, it would appear to be the same path to knowledge of the world that man takes, but there is a big difference: whereas man, through a mechanism of abstraction, seeks a general explanation of the particular phenomena he observes, the agent, in his search, focuses solely on that particular phenomenon and does not go in search of a universal law', thus on the point, G. ACQUISTO, *L'Intelligenza artificiale*, 126 s.

⁷⁸ Just to briefly mention the subject, there are three different modes of learning envisaged for a rational agent: supervised learning, reinforcement learning and unsupervised learning; all three methods respond to different logics. Indeed, in supervised learning, the agent has all the data that characterises a phenomenon and is already broken down into input data and auto-output data and then adapts a certain function to that data that is capable of explaining the phenomenon; once the law has been identified, it can be impeached to predict the output generated by new inputs. Two sets of algorithms fall into this category: regression algorithms and classification algorithms. As for the second category, reinforced learning responds to another learning need that possesses a more dynamic nature. In this scheme, it is assumed that the agent receives sequentially the data it deduces from its observation of the world, and that for each piece of data, it is able to take the action that best enables it to approach a certain known objective function. Specifically, the agent interrogates the world, receiving measurable responses from which it deduces a metric of progress towards the final objective. The dialogue then takes place between agent and world in the form of so-called rewards that the agent obtains for each specific action. Lastly, in unsupervised learning, the agent has no constraints, but has, once again, all the data, as in the case of supervised learning, but cannot distinguish which is the input and which the output of the phenomenon. The aim of this scheme in particular is to identify the underlying law that may have generated such data to be used for future observations, but as no clear distinction is available between different input and output quantities, the result of unsupervised learning ends up being the creation of groupings between data on the basis of a similarity or proximity criterion (this operation takes the specific name of clustering). Take, as an example, the one referred to by several authors; reference is made, in particular, to the artificial intelligence system, built by IBM, capable of playing chess - called Deep Blue -, which defeated the world champion Kasparov in a game developed over six challenges.

On closer inspection, recalling the distinction described in the introduction to this chapter, Artificial Intelligence in its so-called 'strong version', is aimed at achieving and reaching a level of intelligence equal to that of man⁷⁹ (they are, in fact, also defined as 'human-like'⁸⁰). On the contrary, in the perspective of 'weak' Artificial Intelligence, machines behave as if they were actually thinking, capable of cooperating and competing with humans in various (even recreational) activities⁸¹.

Over time, technology has been developing various paradigms that take on increasingly sophisticated conformations and structures in an attempt to improve and enhance computational capabilities. Algorithms intervene in these systems as mathematical tools that are able to extrapolate so-called 'regularities' from a set of data that are encoded in such a way that they can be processed through mathematical formulae⁸².

Secondly, deep learning systems are emphasised in addition to the three original schemes of machine learning (supervised, reinforced, and unsupervised learning).

This is a specific area of machine learning which is based on learning data on different levels, in which the agent performs actions with a higher level of abstraction, typical of humans, such as 'context detection from observing data. Deep learning is realised through the use of a particular way of processing data, a so-called neural network, which is particularly effective today due to the large availability of data and the high computational capacity of machines. Such systems are characterised by the creation of a machine learning model built on several layers. The idea of the 'depth' of deep learning systems is given by the fact that the different learning states or levels are intended to trace the different layers and learning steps of the mammalian brain. On closer inspection, each level of learning should correspond to different areas of the cerebral cortex. In particular, deep learning systems represent a machine learning methodology whereby a system recognises patterns in data through machine learning on a hierarchy of characters. Machine learning (or also deep learning) allows systems to learn

⁷⁹ Take, as an example, the one referred to by several authors; reference is made, in particular, to the artificial intelligence system, built by IBM, capable of playing chess - called Deep Blue -, which defeated the world champion Kasparov in a game developed over six challenges. On this point, he refers to C. PIERGALLINI, *Intelligenza artificiale*, 1745.

⁸⁰ This definition was advocated by A. TURING, *Computing Machinery and Intelligence*, in *Mind*, LIX, 1950, 433 ss.

⁸¹ On this point, please refer to C. PIERGALLINI, *Intelligenza artificiale*, 1745.

⁸² C. CASTELLI – D. PIANA, *Giustizia predittiva*, 20. The authors provide a definition of algorithm extracted again from the CEPEJ works: 'An algorithm is a finite sequence of formal rules (logical operations and instructions), which make it possible to achieve a result from an initial set of input information. This sequence can be part of an automated process and be based on models designed through a process of machine learning', CEPEJ, Ethical Charter for the Use of Artificial Intelligence in Justice Systems and their Environment, 2018.

patterns from the input data and the training received, and to improve them over time, through the acquisition of experience⁸³.

The major difference that distinguishes *machine learning* systems from deep learning systems lies in the different 'depth' and different inner workings. On closer inspection, in *deep learning*, feature extraction from the raw data takes place autonomously, without requiring any 'prior learning' process⁸⁴.

Lastly, it is possible to introduce a concept that will be at the heart of this work: the risk analysis.

This is because, as will be seen in the following paragraphs⁸⁵, when we will introduce the concept of prediction in more detail, we can see how, especially in recent years, the need for predictive analyses or risk assessment tools has become increasingly widespread; in this context, predictive functions have increasingly been entrusted to systems based on automatic training. In this case, reference is made to models (including machine learning) that are built automatically and are able to link the values of predictors to the target to be predicted⁸⁶. Indeed, today, thanks to the integration of available computer resources, advanced artificial intelligence techniques, huge masses of data and great computing power, it is possible to base automatic predictions and evaluations on large sets of examples, each of which can include detailed information.

On closer inspection, Artificial Intelligence, and in particular the systems introduced in this paragraph of machine learning and deep learning, have proved to be extremely useful and indispensable in improving efficiency in the processing of a large amount of data in making prognostic assessments and analysing various risks. In fact, these tools are able to carry out the work of collecting, skimming and entering data, which very often takes a very long time. Such timeframes also prove to be inadequate for the new and emerging needs of various fields of development. Within this framework, it is precisely machine learning tools that reveal their great potential in automating low added-work tasks.

In conclusion, it is inevitable that the A.I. will not be able to "reason autonomously and independently of man"⁸⁷ because the A.I., inevitably, will not be able to possess a total and

⁸³ G. PASCERI, *Intelligenza artificiale, Algoritmo e Machine learning. La responsabilità del medico e dell'amministrazione sanitaria*, Milan, 2021, 23.

⁸⁴ Cfr. The process does not require any reference model to explain to the system the relationship between the different input data. This is possible thanks to artificial neural networks which, as the name suggests, are based on the biological neural network model.

⁸⁵ *Segue*, Chapter 4.

⁸⁶ For example, in the medical field, the learned model can then link symptoms with probable pathologies.

⁸⁷ S. RODOTÀ - E. R. CAPURRO, *European group on ethics in science and new technologies*, Ethical Aspects of ICT Implantés, Human Body, Bruxelles, 2005, No. 20.

"decision-making autonomy and total independence from man" since it will not be able to develop new functions in a conscious manner, since the "biological naturalism" necessary to "substantiate consciousness, "as an emergent phenomenon of the living organism, whose information is encoded in the genome and genetic material of the organism itself" would always be missing⁸⁸.

However, as all instruments of technology, Artificial Intelligence also has inherent limits to its functioning. Indeed, it is incapable of adapting its functioning outside of its own model⁸⁹.

It is precisely in an attempt to question the limits of Artificial Intelligence, while maintaining a purely legal and non-technical approach, that we consider it useful to reflect on the fact that the real world is complex and continuous in mathematical logic; this element already represents a limitation in itself for Artificial Intelligence, which can attempt to identify implicit rules, but would not be able to significantly isolate all the causal factors or modify the models for which it was designed. However, the element that characterises machine learning systems, and at the same time data science, would make it possible to perform at a higher level than human beings (just think of the calculations of large quantities or quantitative analysis); however, they do not provide a solution that is prepared to detect significant qualitative indicators or build reliable projection models in the (imperfect) field of jurisprudence⁹⁰.

3 *Predictive analysis*

⁸⁸ See, E. MAYR-TOWAR, *A New Philosophy of Biology*, Cambridge, 1968, 2.

⁸⁹ In particular, 'The world of the game of Go is defined and definable for a machine: black or white stones, 19 out of 19 lines, exponential combinations of movement, but without the possibility of a complete paradigm shift. Scientists are able to model this kind of world using discrete mathematics, which deals with numerable or finite sets, where millions of trials can be performed to handle the millions of possible combinations. Admittedly, mathematics has taught us that some random events can be modelled, as in Galton's Machine, where a large number of balls fall vertically through rows of nails before settling to the bottom, forming stacks. Although the balls move randomly to the right or left, at the end of the experiment, the heights of these stacks roughly take the shape of a bell curve. But what would happen if there was a total paradigm shift in this model?', we refer to Y. MENECEUR-C. BARBARO, *Intelligenza artificiale e memoria della giustizia: il grande malinteso. Interrogativi su una memoria della giustizia catturata nelle correlazioni dell'intelligenza artificiale*, in *Quest. Giust.*, 2020.

⁹⁰ Today, ICT systems are supporting the massive 'modernisation' and making many human activities more 'automatic' through the 'processing' of a lot of data and information. Modern ICT systems can achieve automated determinations through the use of various approaches and techniques. In particular, reference is made to classical procedural software programmes, developed with traditional procedural programming languages; systems of human-created and formally specified rules; and data mining systems using machine learning algorithms. Data mining is the computational process of discovering patterns in large data sets, using machine learning algorithms and statistics', G. CONTISSA, *Information technology*, 2017, 106.

With the advent and evolution of such artificial intelligence systems we begin to speak of *risk analysis*, which introduces the new concept of *predictive analysis*.

In an attempt to give a definition to a concept which it can be used in so many fields and other sectors, predictive analysis consists of '*using data, statistical algorithms and machine learning techniques to identify the probability of future outcomes based on historical data*'. In other words, indeed, the panorama widens further as the concept itself encompasses, on closer inspection, three macro factors: predictive modelling, machine learning and data mining⁹¹.

It uses *predictive analytics*, i.e. an advanced tool capable of extracting values from big data⁹². These tools process huge amounts of data and, with the support of artificial intelligence, predictive analytics is able to provide a prediction of the future⁹³. This type of analysis, linked to the result, has undoubtedly proved useful in many fields⁹⁴ and applications, including

⁹¹ There are different types of analysis, descriptive, prescriptive and predictive. They all possess the same data analysis structure but are directed in different directions. "Data mining attempts to identify inherent patterns within large amounts of data by using mathematical and stochastic processes and algorithms. At best, from the results obtained in this way, it is possible to read and anticipate *trends* and potential developments.

⁹² The magnitude evoked by the term 'big' has at least two meanings, referring not only to the amount of data processed, but also to the extraordinarily deep scope and granularity of the analyses that can be performed on them', thus on the point S. FARO-T. E. FROSINI - G. PERUGINELLI, *Dati e algoritmi. Diritto e diritti nella società digitale*, Bononia, 2020.

⁹³ Consider, for example, the protection in the field of external state security. In fact, especially in recent years, the so-called 'no-fly lists' have become widespread, special lists that make it possible to foresee and plan security checks, which, however, have generated various issues that have concerned the current debate: discriminatory problems, for example, linked to the membership of a particular ethnic group. If one broadens the panorama overseas, one thinks of the example of a small town in the United States, in Memphis, which, by analysing the frequency of crime locations and times, crime was reduced by around 25% because the algorithm used allowed the police to predict the places where crimes would be committed.

⁹⁴ Think also of the health system and some peculiar impacts on the population: in fact, by analysing searches by Google users on the symptoms of certain diseases, it was possible to predict the spread of an epidemic. The project in question is called Glu trends. On this point, see M. MATTIOLI, *Disclosing Big Data*, in *Minn. L. Rev.*, 2014, 535. In addition, it is also possible to manage hospital waiting lists more efficiently. For example, and in this regard, from Internet searches for two drugs in relation to hypoglycaemia, researchers hypothesised that there might be a correlation between the use of the two drugs and a side effect, and experimental tests have since confirmed the hypothesis. C. ANDERSON, *The end of theory: the data deluge makes the scientific method obsolete*, in *wired.com*, 23 June, 2008. The fields of application are boundless, one thinks, for example, even in the university sphere of planning or the VWR or VTR evaluations that are used by universities to put a post up for competition or not; or the distribution of FFO among universities or the medians that also allow one to become an ASN commissioner or to participate in a habilitation procedure. There is also the subject of spatial planning: in fact, spatial analyses allow urban planning or traffic management as well. Think also of utility management: predicting the load on the electricity grid is an example of predictive analysis from which decisions on how to use power plants are derived. On this point, please refer to an interesting analysis by M. Paterson-M. McDonagh, *Data protection in an Era of big data: the challenge sposed by big personal data*, 2918, in *Monash U. L. Rev.* 1, p. 6. Another and last case in Italy refer to their use in the social security system: the INPS case in which the big data system has been implemented since 2000 to support the Institute's strategic decisions. The system also acquires information from supervisory inspectors that is useful in particular for 'anti-fraud intelligence' activities that detect cases of double payment of family allowances within the same family, of undue adjustment of sickness benefits. Lastly, even in environmental matters where the right of access finds its fullest expression: in 2015, the Best Policy Insights Hack prize was awarded to a private software programme that indicated, thanks to big data, in which neighbourhoods it was preferable to install solar panels. On this point, see a F. COSTANTINO, *Intelligenza artificiale e decisioni amministrative*, in *Riv.it.sc.giur.*, 2017, 370.

industry⁹⁵. In particular, predictive analysis uses the tools discussed in the previous paragraph ;it consists of using data, statistical algorithms and artificial intelligence and machine learning techniques in order to identify the probability of future outcomes, based on historical data. In particular, it is an extremely advanced form of *Business Intelligence*.

Nevertheless, it seems appropriate to point out why this analysis is relevant and finds its point of encounter with criminal law⁹⁶. In this regard, at this first and introductory stage, it will suffice, while maintaining an approach to the criminal law system, to turn our attention to certain instruments which, to date, are considered, due to the considerable and lively debate surrounding them, to be new protagonists within the criminal law panorama⁹⁷.

3.1 *The new players in analytical prediction: predictive algorithms*

In the new technological scenario, human activities have evolved and improved to the point where they are confronted with new tools called predictive algorithms.

First, it is necessary to introduce this concept and try to understand its essential characteristics. Indeed, the term 'algorithm' is often used to refer to A.I. applications⁹⁸. These are true procedures that are also susceptible to automatic application, having a field of use that can extend beyond artificial intelligence systems, but encompassing all types of computer systems (as they can also possess varying degrees of complexity)⁹⁹.

⁹⁵ "For companies facing a market that is more and more competitive every day, Predictive Analytics is increasingly proving to be a decisive tool, because it allows them to identify patterns and trends and obtain estimates and anticipations of how they will evolve", P. LICATA, *Predictive Analytics*, in *Digital4 Online*, 30 March 2022.

⁹⁶ "Can criminal law play a role in this? If so, what kind? The unease at the prospect of an answer is palpable. In the face of AI, criminal law, needless to conceal it, risks appearing to be an old tool that has always been tailored to man in the flesh: the algorithm (the new 'mind') releases a different and unfathomable (at least for now) hermeneutic, so much so that one would be inclined to wave, suitably readjusted, the motto "*Silete poenologi in munere alieno!*" And yet, the problem of 'regulation' and 'liability' has attracted the attention of criminal and procedural penal science", thus on the point C. PIERGALLINI, *Intelligenza artificiale*, op. cit., 1746 s. The author, in the main introduction, questions what precisely the meeting point might be, but above all the role of criminal law as it relates to new technological tools.

⁹⁷ Please refer to the following section for a more detailed discussion on this subject, § 3.1.

⁹⁸ It is no coincidence that it is often found when one wants to refer, even loosely, to 'decision-making systems'. It is also mentioned when, especially in recent times, one speaks of 'algorithmic governance'. The risk today is that of making a big cauldron in which, all the issues related to A.I. are immediately reconnected to the algorithmic concept and tool, even though they are only one aspect of it. Issue raised by G. SARTOR, *L'intelligenza artificiale e il diritto*, 9.

⁹⁹ Moreover, "AI algorithms perform various epistemic and practical functions (related to reasoning, perception, classification, planning, decision-making, etc.). Some algorithms merely apply pre-existing knowledge, others perform forms of learning, helping to create or modify the model on which the functioning of the system of which they are part is based. For example, an AI system for e-commerce could limit itself to applying predetermined rules (e.g. applying discounts to consumers who meet certain conditions) but could also learn and use correlations between users' characteristics and activities and their preferences (to recommend purchases) and develop and select effective strategies for commercial activity (to negotiate online, or optimise financial management), see again G. SARTOR, *L'intelligenza artificiale e il diritto*, 10.

It should be pointed out from the outset that there are, to date, enthusiastic attitudes on the subject, fascinated by these tools but, at the same time, there is no shortage of critical voices¹⁰⁰ that analyse, according to a problematic reading, the prospects related to their use in the awareness of the advantages that may derive from an *evidence-based* ascertainment system, on the one hand, and the risks to the protection of fundamental rights, on the other. The term 'algorithm', which is related to the term 'society', can then take on a variety of meanings depending on the context in which it is used, with nuances and variations that are often not shared by experts in the same field. One accredited definition, however, defines them as “*algorithms need not be software: in the broadest sense, they are encoded procedures for transforming input data into a desired output, based on specified calculations. The procedures name both a problem and the steps by which it should be solved*”¹⁰¹. Such 'encoded procedures' presuppose the realisation and use of a computational model, which reproduces a phenomenon, taking all relevant variables into account and regulating their interaction.

Indeed, it should be noted from the outset, how they have timidly approached the criminal justice system. The explanation for this approach can undoubtedly be found in the fact that they appear and acquire a great deal of fascination in the face of a system which - due to its structure and present institutions - in many cases (even in the decision-making phase) requires the human being to make prognostic evaluations directed towards the future.

An attempt will be made to provide, firstly, an initial definition of a predictive algorithm, which will serve as a foundation for the analysis to be carried out in the following paragraphs.

Undoubtedly, as will be seen in the following section, in which instead an attempt will be made to provide a more precise picture of the functioning of algorithmic software, these have undoubtedly given rise to various doubts directly connected with their operation, determined

¹⁰⁰ "Are we perhaps at the beginnings of a shocking change in the traditional scenario of criminal jurisdiction, in a profound and restless reshuffling of the typical coordinates of the two paradigms, circumstantial and Galilean, which no longer seem conceptually distinct and autonomous? In the face of the technical complexity and fatigue of traditional judicial operations reconstructing the fact, is post-modernity undermining the fairness, effectiveness and guarantees of the model of critical rationalism, or does the art of judging, albeit 'reasoning under uncertainty' and 'by probabilities', remain firm and vital? What will be the new frontiers of crime control strategies for criminal justice: from fair justice to exact justice?", G. CANZIO, *Intelligenza artificiale, algoritmi e giustizia penale*, in *Sistema penale*, 8 January 2021.

¹⁰¹ For the limited purpose of these brief reflections, we accept the definition offered by Tarleton Gillespie, in 2014, which was also taken as a paradigm by the valuable study already published by the Council of Europe, T. GILLESPIE, *Algorithms and Human Rights* in December 2017. Indeed, see, T. GILLESPIE, *The relevance of Algorithms*, in T. Gillespie - P. Boczkowski - K. Foot (eds), *Media Technologies*, Cambridge US, 2014, 167.

by their influence in the selection of data and in the "construction of correlation relationships between information and the predictions or configuration of probable future scenarios"¹⁰².

The entry of such tools raises several questions, including the need to find new balances capable of reconciling the human need for greater certainty and rapidity in prognostic evaluations and the reliance on technologies purely based on calculations and automated systems. The statistical predictability belonging to humans is being put to the test in a bench that sees the entry of new computational and mathematical powers derived from the new artificial agents on the scene, which in turn are governed by statistical calculations and engineers.

In conclusion of this premise, it is necessary to anticipate that, although artificial prediction is already being used¹⁰³ in the assessment of the risk of criminal recidivism and in sentencing, this does not mean that such use is free of implications and issues that arise and make such tools impartial¹⁰⁴. It will be necessary, in the remainder of the paper, to assess the risks, implications and benefits that may arise, starting from the first application implications of such instruments in a supranational context.

3.2 *Brief outline of how algorithmic software works*

On closer inspection, as already mentioned, the main activity belonging to algorithms is their extraordinary ability to abstract "regularities and patterns from a large amount of information"¹⁰⁵. In an attempt to provide a conceptual definition of them, they can be defined as a "sequence of computational instructions that tell a machine the procedure to follow to obtain a certain result"¹⁰⁶.

Algorithmic operation is mainly based on two main criteria: firstly, the possibility of predicting the repetition in the future of events or facts that have already occurred; the second, on the other hand, concerns more specifically the mechanism of attribution of 'similarity between persons, which establishes a criterion according to which certain similar entities are

¹⁰² U. RUFFOLO, *Intelligenza artificiale*, op. cit. p. 12 s. On this point, see also A. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro della libertà*, in *BioLaw, Journal-Rivista di BioDiritto*, No. 1, 2019.

¹⁰³ *Segue*, Chapter III.

¹⁰⁴ On this point, please refer to V. MORIGNAT, *L'I.A., dalle previsioni alle decisioni*, in A.F. URICCHIO – G. RICCIO – U. RUFFOLO (eds), Bari *L'intelligenza artificiale tra etica e diritti. Prime riflessioni a seguito del libro bianco dell'Unione europea*, 2020, 49.

¹⁰⁵ U. RUFFOLO, *Intelligenza artificiale*, 12 ss.

¹⁰⁶ This is, undoubtedly, a manualistic definition, taken from legal informatics manuals. G. AVANZINI, *Decisioni amministrative e algoritmi informatici. Predefinizione analisi predittiva e nuove forme di intelligibilità*, Naples, 6.

able to behave in a similar manner upon the occurrence of certain circumstances in a given environment or sector'.

On closer inspection, the simultaneous application of both criteria represents a strong tool for increasing and overcoming prejudices and stereotypes, accepted as valid due to the algorithm's supposed neutrality.

Now, the automatic perception of absolute trust of the human being towards algorithms or technological tools, not adequately complied with and balanced, would run the greatest risk of 'marginalising and even discriminating against certain categories of individuals' or, furthermore, that of producing a generalised conformation of citizens to the most frequent behaviours prescribed by algorithmic predictions. Thus, on the one hand, as a consequential effect, the phenomenon of so-called *marginalisation* would be avoided and, at the same time, the space for free choice left to individuals would be reduced.

Taking full awareness of the inexistence of algorithmic objectivity, it appears more than necessary, in the use of these tools, to monitor, with particular attention, the activities and functioning of algorithmic software, used to make, in some cases, decisions, capable of impacting the lives of individuals or the community.

Nevertheless, in an attempt to provide a - albeit brief - background on how algorithmic software works, the analysis and development of the paper cannot be separated from the introduction of the notion of 'big data'.

The term itself already gives the idea that we are dealing with data, and the English-speaking adjective implies the large quantitative dimensions of the same¹⁰⁷. However, there is still no unambiguous definition of the concept of big data, which according to Dumbill, big data is "*data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the strictures of your database architectures. To gain value from this data, you must choose an alternative way to process it*"¹⁰⁸. We are talking about large and massive amounts and sets of electronic data that the various tools of technology (smartphones, computers) and its spread and cost now make accessible to all individuals. Every individual, therefore, is able, consciously, or unconsciously, to produce an indefinite amount

¹⁰⁷ Temporally, one speaks of big data as early as the mid-1990s, and it is from the year 2000 onwards that their definition is more clearly affirmed, which hinges on the four 'Vs' (Volume, referring to the quantitative aspect; 'velocity' referring instead to the speed with which data can be generated or transmitted; 'Variety' referring to the diversity of data and their origin; 'veracity' referring instead to the quality, correctness and reliability of the data. On this point, see also C. COMELLA, *Origine dei "Big data"*, in *Gnosis*, 2017.

¹⁰⁸ E. DUMBILL, *What is big data*, in *Big Data Now: current perspectives*, O mMedia, O' Reilly Media: California, 2012, 3.

of data¹⁰⁹ and to 'leave a digital trace' of all the various actions that he or she performs on a daily basis in the daily street¹¹⁰.

However, what can undoubtedly be noted is that the term big data is normally applied to huge collections of data that are difficult to process using the information technologies that are used for digital systems (those that collect documents, for example).

This peculiar type of data has three main characteristics: volume, high speed (in change) and great Variety¹¹¹. On closer inspection, the data that make up Big Data can be created by humans, although in most cases they are collected automatically by devices capable of capturing and collecting data from the external physical world.

The main feature, or perhaps the one that stands out the most, of this type of data is that the distinctive element that makes a mass of data 'big' is a functional feature: indeed, the possibility of using that data for 'analytics' purposes, i.e. to discover correlations and make predictions. To this end, this is one of the reasons why more and more machine-learning-based A.I. technologies are being used in today's world, which allow predictive models to be extracted from large data sets¹¹².

As a continuation of what was mentioned in the previous section, we see how an A.I. system may comprise several algorithms within it, the interaction of which results in the functioning of the system itself. Indeed, even within an algorithm (which may be simpler or more complex) that includes algorithms that perform specific functions. On closer inspection, it is also noticeable that in systems constructed for them to 'learn', the most important component is not

¹⁰⁹¹⁰⁹ “Big data are usually described and characterized with a reference to the “3Vs”, namely volume, velocity, and variety. Volume, because “volumes of data are larger than those conventional relational database infrastructures can cope with”. Velocity, because “it’s not just the velocity of the incoming data that’s the issue: it’s possible to stream fast-moving data into bulk storage for later batch processing, for example. The importance lies in the speed of the feedback loop, dating data from input through to decision. Variety, because very often data present itself in a form perfectly ordered and ready for processing. Instead, they are unstructured (e.g. text from social networks, image data, a raw feed directly from a sensor source, etc). several enabling factors provided the conditions for the exponential growth of Big Data [...]”, G. CONTISSA, *Information technology for the law*, Turin, 2017, 104; also E. DUMBILL, *What is big data*, in *Big Data Now*, 5.

¹¹⁰ It is therefore no coincidence that one area where the legislator has readily intervened is the regulation of relations between the public administration and big data, since the public administration stands as a key reference point, since it is by far the largest holder of data. In particular, the relations between big data and the administration could be summarised by following three logical stages: the first stage, also called 'digitalisation'; the second that could also be defined as 'data interconnection' and, finally, the stage we are in at the moment which is represented by the application of artificial intelligence to administrative decisions, or also called the 'predictive analysis' stage. Following this line of analysis and reflection, we refer to F. DE LEONARDIS, *Big Data, decisioni amministrative e “povertà” di risorse della Pubblica amministrazione*, in E. Calzolaio (ed), *La decisione nel prisma dell'intelligenza artificiale*, Milan, 2020, 140 s.

¹¹¹ Other characteristics sometimes associated with big data are low Veracity (high probability that some information is inaccurate) and high Value (the usefulness, correlated with the breadth of mass, that can be derived from the data through analysis techniques).

¹¹² G. SARTOR, *L'intelligenza artificiale e il diritto*, 12.

so much the algorithmic model constructed in part by the system in order to be able to perform the given tasks. On the contrary, the core of the system is rather the learning algorithm that generates the algorithmic model, based on the data to which the system has access¹¹³.

In recent years there has been a *dramatic growth* in the amount of data expressed in digital form¹¹⁴. For this reason, it is necessary to introduce, albeit briefly, the concept of big data, in order to arrive at the deeper considerations concerning algorithmic functioning. In particular, the algorithm is the tool that is able to 'make the data speak'¹¹⁵ through the search between different correlations¹¹⁶. More specifically, we are talking about the so-called predictive and recommendation algorithms that 'enable predictions of future behaviour to be made, starting from the analysis of past behaviour that is well known to companies and private operators involved in marketing'¹¹⁷.

The ability to process data in real time has stimulated the collection/collection of huge amounts of data. These new software tools are generally based on Artificial Intelligence approaches.

And it is precisely through the various interconnections and processing of data that knowledge of phenomena and their trends is achieved and, consequently, the so-called government by data. Therefore, one can see at first glance that the great capacity and point of favour of using such tools is undoubtedly represented by the possibility of exponentially expanding the knowledge base of data, which was unimaginable until a few years ago, but

¹¹³ As an example, in a classifier system that recognises images by means of a neural network, the crucial element is not the neural network, but rather the learning algorithm (the 'learner' algorithm) that modifies the structure of the neural network (the algorithmic model) by changing the weights of its connections, so that it improves its performance in classifying objects of interest (e.g. animals, sounds, faces, attitudes, feelings, etc.), thus, G. SARTOR, *L'intelligenza artificiale e il diritto*, 10.

¹¹⁴ In accordance to L. FLORIDI, "*In the 2003, researchers at Berkeley's School of Information Management and Systems estimated that humanity had accumulated approximately 12 exabytes of data (1 exabyte corresponds to 1018 bytes or a 50,000-year-long video of DVD quality) in the course of its entire history until the commodification of computers. However, they also calculated that print, film, magneti, and optical storage media had already produced more than 5 exabytes of data just in 2002*", L. FLORIDI, *Information: A very short introduction*, Oxford, 2010, 6.

¹¹⁵ F. DE LEONARDIS, *Big data*, 141.

¹¹⁶ On this point, see also G. AVANZINI, *Decisioni amministrative e algoritmi informatici*, 11.

¹¹⁷ The Italian legal system, in this regard, contains two fundamental provisions: firstly, Article 50 CAD stipulates that public administrations, within the scope of their institutional functions, proceed to the analysis of their own data, also in combination with other administrations, and Article 50 ter, on the other hand, which institutionalises a national digital data platform project (the PDND). There is also a three-year plan concerning information technology born within the public administration, covering the three-year period 2017-2019, which deals with artificial intelligence even though it postpones its implementation precisely to the realisation of the structure considered indispensable for the exploitation of big data, the National Digital Data Platform (PDND), which is entrusted with the task of enhancing the information assets of the public administration through the use of big data technologies at the service of public decision-makers, thanks to exploratory data analysis. Also of fundamental importance is the White Paper that was adopted for the Future of Digital Italy (AGID), dedicated to artificial intelligence published in 2018 and which considers this institution as a tool for social, economic and cultural development.

another factor, which is no less important, is represented by the possibility of such tools to correlate data and generate new bases and new fields of knowledge that, until then, for the human mind, were limited to factual knowledge and connections to which man could be pushed.

4 *The “shy entry” of Artificial Intelligence into the courtrooms*

Technological advances of algorithms and machine learning now also affect law.

They introduce the use of tools with very high computing power, capable of processing huge amounts of data with increasing efficiency and autonomy: from the most advanced jurisprudential search engines to computer tools for the drafting of documents, from 'algorithmic proofs' to software capable of processing enormous amounts of data with ever increasing efficiency and autonomy algorithmic evidence' to software capable of predicting the outcome of a trial or resolving a dispute; in this sense artificial intelligence seems potentially able to penetrate every aspect of judicial activity. The question arises as to the possible advantages that these new technologies could bring to the justice system, starting from the reduction of trial times and the decrease in the possibility of error, passing through the benefits in terms of the deflation of litigation (which would be partially replaced by predictive mechanisms and computerised alternative dispute resolution systems) and of greater territorial uniformity between court decisions¹¹⁸.

To date, it should be premised that Artificial Intelligence approaches justice as a mere support tool and not with the idea of completely replacing human activities. There are, several

¹¹⁸ A. PAJNO and others, *A.I.: profili giuridici. Intelligenza artificiale: criticità emergenti e sfide per il giurista*, in *Biolaw Journal*, no. 3/2019, 226-227; for an overview of justice efficiency problems, also C. CASTELLI-D. PIANA, *Giusto Processo e Intelligenza Artificiale*, 25-46.

application processes in various justice activities concerning predictive justice¹¹⁹ mechanisms in our country that have not yet found application¹²⁰.

In this regard, the doctrine has circumscribed the scope of the investigation, identifying four main scenarios within which 'the technological revolution set in motion by the I.A. could most significantly impact the claims of protection of legal assets'¹²¹.

Firstly, A.I. tools find favorable application in law enforcement activities¹²² and, in particular, in predictive policing¹²³. In this case, A.I. systems and, in particular, predictive algorithms, could undoubtedly make an important contribution to preventing the commission of crimes and arranging for a more efficient allocation of resources. In this field, there are several software programs that have been implemented in various Italian police headquarters

¹¹⁹ On this point, an interesting analysis on the risks and drifts of predictive justice, S. QUATTROCCOLO, *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale "predittiva"*, in *Cass. Pen.* No. 4, 2019, p. 1748; ID., *Intelligenza artificiale e giustizia: nella cornice della Carta etica europea gli spunti per un'urgente discussione tra scienze penali e informatiche*, in *Leg. Pen.*, 18 dicembre 2018; ID., *Processo penale e rivoluzione digitale: da ossimoro a endiadi*, in *medialaws*, 2020, 3, 1 ss.; Among the numerous studies on the subject of predictive justice, see in particular F. BASILE, *Intelligenza artificiale e diritto penale*, cit.; C. CASTELLI-D. PIANA, *Giustizia predittiva. la qualità della giustizia in due tempi*, in *Quest. Giust.*, No. 4, 2018, 154 ss.; ID., *Giusto processo e intelligenza artificiale*, Rimini, 2019; V. MAFFEO, *Giustizia predittiva e principi costituzionali*, in *www.i-lex.it*, 2019, 12, 277 ss.; E. NAGNI, *Artificial intelligence, l'innovativo rapporto di (in)compatibilità tra machina sapiens e processo penale*, in *Sist. Pen.*, No. 7, 2021, 5 ss.; G. RICCIO, *Ragionando su intelligenza artificiale e processo penale*, in *Arch. Pen.*, No. 3, 2019, 1 ss.; E. RULLI, *Giustizia predittiva, intelligenza artificiale e modelli probabilistici. Chi ha paura degli algoritmi?* in *Analisi giuridica dell'economia*, 2018, 537, ss.; E. STRADELLA, *La regolazione della Robotica e dell'Intelligenza artificiale: il dibattito, le proposte, le prospettive. Alcuni spunti di riflessione*, in *www.medialaws*, 2019, 1, 73 ss.; G. TAMBURINI, *Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiale*, Roma, 2020; C. TRAVISI, *La regolamentazione in materia di intelligenza artificiale*, in *www.medialaws.eu*, No. 2/2018; R. TREZZA, *Diritto e intelligenza artificiale, Etica, Privacy, Responsabilità, Decisione*, Pisa, 2020.

¹²⁰ On the other hand, the case of the policing tools used by several Italian public prosecutors' offices aimed at crime detection and better resource efficiency is different. However, the debate on the subject is very heated and it certainly seems useful to provide an insight into which scenarios are most discussed today. Per un completo inquadramento della materia della predictive policing, see W.L. PERRY - B. MCINNIS - C.C. PRICE - S.C. SMITH - J.S. HOLLYWOOD, *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*, Rand Corporation, 2013.

¹²¹ F. BASILE, *Intelligenza artificiale e diritto penale*, 16.

¹²² G. ITALIANO, *Intelligenza artificiale, che errore lasciarla agli informatici*, in *Agendadigitale.eu*, 11 June 2019, which notes that "A.I. techniques are and can be successfully employed in the analysis of available information, transactions, log files, network traffic, and all the 'footprints' that each individual leaves on the network and in digital systems, in order to identify possible anomalies and suspicious activities, or simply to compose in a coherent vision the information coming from multiple and heterogeneous sources, and extract knowledge from it, so as to automatically make decisions or provide support to human decision-makers, who must be able to react faster and faster to external stimuli".

¹²³ "Predictive policing" can be understood as the set of activities aimed at the study and application of statistical methods with the objective of 'predicting' who may commit a crime, or where and when a crime may be committed, in order to prevent the commission of crimes. Prediction is fundamentally based on an actuarial reworking of different types of data, including reports of previously committed crimes, the movements and activities of suspects, the locations of recurrent criminal acts, and the characteristics of these locations, the time of year or the weather conditions most likely to be associated with the commission of certain crimes; among the data used for these purposes, information on ethnic origin, level of schooling, economic conditions, and somatic characteristics sometimes also appears (...). Lombroso's revenge?), ascribable to persons belonging to certain criminological categories (e.g., potential terrorists), etc. [...], *Cfr.* F. BASILE, *Intelligenza artificiale e diritto penale*, 10.

and that to date, are bringing positive results, succeeding in reducing the commission of certain crimes¹²⁴.

Secondly, a further application viewed favourably by commentators, is found in the opportunity of employing decision-making algorithms to resolve criminal disputes; in particular, such predictive algorithms would act as a means of flanking the judge-man with the judge-machine¹²⁵.

The third field of investigation which has most interested the attention of Italian commentators concerns the use of such tools in the assessment of criminal assessment of dangerousness. On closer inspection, the idea has begun to be put forward, albeit at an embryonic and merely theoretical level, of entrusting such an assessment to predictive algorithms, capable of drawing on and processing enormous quantities of data in order to bring out relations, coincidences and correlations, which would make it possible to profile a subject and foresee his subsequent behaviour, even of criminal relevance¹²⁶. In Italy, given the extensive use of certain instruments such as security and prevention measures, the debate on the subject is finding fertile ground, since there are several moments when the judge is asked

¹²⁴ On the subject of policing systems in Italy, see F. BASILE, *Intelligenza artificiale e diritto penale*, 11 ss. In particular, on the possible uses of such instruments, the author reports that "Moreover, the results provided by this software could in some cases be used not only for predictive purposes, but also to reconstruct the criminal career of the profiled subject, i.e. to have an investigative trail to follow in order to charge him/her not only with the last crime committed (on the occasion of which he/she was detected), but also with the previous crimes constituting the criminal series reconstructed thanks to the storage and processing of the data"; A.D. SIGNORELLI, *Il software italiano che ha cambiato il mondo della polizia predittiva*, in *Wired.it*, 18 May 2019; C. PARODI-V. SELLAROLI, *Sistema penale e intelligenza artificiale: molte speranze e qualche equivoco*, in *Riv. Trim. - Dir. Pen. Cont.* no. 6/2019, 56. For a description of Keycrime, provided by its creator, Mario Venturi, see ID., *La chiave del crimine*, in *Profiling*, 4, 2014. On the initial concerns arising from the use of such instruments, see R. PELLICCIA, *Polizia predittiva*.

¹²⁵ On the initial questions that arose in the margins of the possibility of using such instruments in decision-making, see A. TRAVERSI, *Intelligenza artificiale applicata alla giustizia*, cit., 3. On the possibility of applying these instruments in the decisional phase in order to overcome the scrutiny envisaged by Article 533(1) of the Code of Criminal Procedure, see a G. CANZIO, *Il dubbio e la legge*, in *Diritto penale contemporaneo*, 2018, 1 ss.; M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, in *ivi*, 2019, 1 ss.; A. NATALE, *Introduzione. Una giustizia (im)prevedibile?*, in *Questione Giustizia*, No. 4, 2018, 1 ss.; in the same dossier, see also the contributions of C. COSTANZI, *La matematica del processo: oltre le colonne d'Ercole della giustizia penale*, e di C. CASTELLI- D. PIANA, *Giustizia predittiva. La qualità della giustizia in due tempi*. See, finally, the monographic issue of *Giurisprudenza Italiana* for the hundred and seventy years of the journal, dedicated to the topic of Artificial Intelligence (forthcoming).

¹²⁶ In recent years, the debate in Italy has also been focusing on the possibility of introducing an evidence-based approach to assessing an individual's risk of future dangerousness. On this point, an interesting analysis by G. ZARA, *Tra il probabile e il certo. La valutazione dei rischi di violenza e di recidiva criminale*, in *Diritto penale contemporaneo*, 20 May 2016. On the factors considered in this type of evaluation issued by the algorithm, please refer to L. CASTELLETTI- G. RIVELLINI- E. STRATICÒ, *Efficacia predittiva degli strumenti di Violence Risk Assessment e possibili ambiti applicativi nella psichiatria forense e generale italiana*, in *Journal of Psychopathology*, 2014, 153 ss.; G. ROCCA- C. CANDELLI- I. ROSSETTO- F. CARABELLESE, *La valutazione psichiatrico forense della pericolosità sociale del sofferente psichico autore di reato: nuove prospettive tra indagine clinica e sistemi attuariali*, in *Rivista Italiana di Medicina Legale* (e del Diritto in campo sanitario), No. 4, 2012, 1442 ss.

to make prognostic evaluations of an individual¹²⁷. Finally, the possible hypotheses of involvement - as an instrument, as a perpetrator, or as a victim - of an A.I. system in the commission of an offence are also assessed¹²⁸.

Secondly, a number of experiments have been conducted in the Italian judicial system in recent years in order to develop 'Predictive Jurisprudence' systems¹²⁹.

In 2018, the Court of Appeal and the Court of Brescia also launched an experiment in 'predictive justice' through a project aimed at predicting the length of a case and the principles to be adopted when assessing it. Such a system will, in the best-case scenario, also make it possible to estimate whether or not a claim can be approved. Such systems are, today, being tested not only in courtrooms, but also in the private sector, within law firms. The analysis of court decisions by means of artificial intelligence tools allows lawyers to know in advance the likelihood of success in a given case¹³⁰.

Lastly, we would like to mention the *CrossJustice* project, set up within the University of Bologna, which aims to develop an online platform for advice and support on the effectiveness of procedural rights that provides a free service, directed mainly at legal professionals, but accessible to law students, NGOs and all European citizens¹³¹.

¹²⁷There is an extensive debate on the subject in Italian doctrine. On this point, reference is made in particular to the analysis of the different areas within which it is necessary to formulate a *prognosis* of a future offence to F. BASILE, *Esiste una nozione ontologicamente unitaria di pericolosità sociale? Spunti di riflessione, con particolare riguardo alle misure di sicurezza e alle misure di prevenzione*, in *Riv. it. dir. proc. pen.* 2018, 644 s.

¹²⁸F. BASILE, *Intelligenza artificiale e diritto penale*, *ivi*, 2 ss.

¹²⁹In 2019, a first experiment was developed by the LIDER Lab of the Scuola Superiore Sant'Anna, in Pisa, in collaboration with EMbeDS, KDD Lab and the Court of Genoa. The system developed by the researchers makes it possible to have access to all the sentences issued by the Court itself and to analyse them using complex machine learning mechanics. The analysis conducted by the A.I. then allows the judge using it to identify guidelines, common trends and case law practices in relation to each specific case under examination. This would then allow the judge to more accurately verify the consistency of his or her own position with that expressed by other colleagues in similar cases. The project follows five autonomous and interconnected levels: the first level concerns the analysis of judgments and related court documents, according to the criteria and methodologies developed in the Observatory on Personal Injuries, applicable to areas of litigation other than those concerning non-monetary damages; the second and third levels concern the analysis of the same data by means of machine learning techniques in order to develop both tools for the annotation and automatic extraction of information from legal texts (level 2) and algorithms for analysis and prediction (so-called 'level 3 artificial intelligence'); The database will be constructed in such a way as to allow the development of algorithms that identify trends according to criteria known to the user agent, and at the same time highlight new trends on the basis of possible biases and trends discovered by the algorithm itself. The fourth level concerns the understanding of the rationale behind each decision and the development of suitable tools to explain the criteria defined by the artificial intelligence; the last concerns the structuring of an analysis of the legal argument at a level of abstraction and systematicity useful for the simplification of all tasks.

¹³⁰On possible future scenarios, see M. GIALUZ, *Quando la giustizia penale incontra il diritto penale: luce e ombre dei risk assessment tools tra Stati Uniti e Europa*, in *Arch. Pen.*, 29 May 2019, 19 ss.

¹³¹The *CrossJustice* platform provides an innovative architecture with the aim of providing support for: 1. the conformity of national instruments implementing EU directives with the EU acquis; 2. the compatibility of national frameworks resulting from the implementation of EU directives. Thanks to this set of resources and functionalities, the *CrossJustice* platform will contribute to meeting various needs in the field of criminal procedural law. In particular, it will contribute to: provide ICT-supported analysis and evaluation of the

The approach of these instruments and their possible entry into the penal system, infiltrates and branches out in different directions: firstly, the most evident clash occurs if one looks at the fact that criminal law is purely and centrally 'built on man, on personal and guilty reproach, on the degree of responsibility and reproachability for a human action'¹³². In particular, human judgement and the attribution of responsibility to a subject under judgement, is an action that is purely and simply left to man who, with his 'human capacities and limits of evaluation, discernment, and criticism', finds himself daily generating judgements attesting to the responsibility or otherwise of a given subject, but also the decision on the application of other peculiar types of measures, such as, for example, security or prevention measures.

Well then, in the face of a legal system that requires a human being to make a decision that, although anchored to factual elements, possesses within itself a prognostic assessment that looks to the future, in this precisely field predictive algorithms or artificial intelligence tools gain most acceptance.

The analysis of the paper, also for systematic reasons, will be aimed at following two main macro-guidelines: the prognostic analysis of dangerousness carried out by a male judge and supported by an A.I. tool and, secondly, a very interesting aspect, the use of such tools as a support to the judge in choosing the best sanctioning treatment¹³³, two aspects both temporally and substantially, undoubtedly linked.

4.1 The first automated decisions in civil and administrative sectors: the disruptive technologies

On closer inspection, some Artificial Intelligence tools have already been applied in some States¹³⁴ in order to solve certain civil and administrative disputes. In an attempt to provide a general overview of these type of applications, for methodical and systematic reasons, in this thesis I selected only a few of them¹³⁵.

compliance of national legislation with the relevant EU acquis; and support capacity building of legal practitioners, enabling lawyers and students to complete their knowledge of the national implementation of the EU acquis on procedural rights.

¹³² Thus, authoritatively V. MANES, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in U. Ruffolo (ed), *Intelligenza artificiale. Il diritto, i diritti e l'etica*, Milan, 2020, 547.

¹³³ The two macro-guidelines addressed by the paper will be explicitly dealt with in Chapters II and III.

¹³⁴ Consider the case of Estonia, which is the first country in Europe to rely on artificial intelligence tools in some areas of justice. The ultimate aim is to lighten the work of judges and chancelleries not with a gigantic amnesty or a series of 'simplified' judgments or voluntary judges, but by using an artificial intelligence system to relieve the tasks of chancelleries and magistrates. The service will be launched later this year and will cover civil cases with a maximum value of seven thousand euro.

¹³⁵ For example, in Italy is one of those countries where the use of an algorithm has found application in certain administrative proceedings. On the current applications in Italy and the first jurisprudential implications, see the next paragraph and the considerations in the margin of the first relevant decisions.

In Europe, the use of artificial intelligence as a work support tool for legal practitioners and judges is still an embryonic phenomenon. It should be noted that in civil and administrative matters, the disruptive approach of Artificial Intelligence has different effects and repercussions than in the criminal justice system.

4.1.1 Italian jurisprudence and the A.I.

In Italy, the administrative judge has been faced with the process of automation in administrative proceedings. There have been some important recent judicial decisions that have intervened on the use of these instruments in administrative proceedings¹³⁶.

On closer inspection, it is interesting to note the position taken by the jurisprudence of the Council of State, which has been dealing extensively with the use of automatic mechanisms based on artificial intelligence since as early as 2019¹³⁷, albeit in the exclusive context of administrative procedures¹³⁸. The strong point of this decision, which in some ways represented the 'official' opening of justice with regard to the I.A., in general, is the fact that it was established that 'the use of computerised procedures cannot be a reason for circumventing the principles that shape our system'. On closer inspection, this is a strong stance that undoubtedly shows, firstly, the cautious attitude of a legal body and, secondly, the necessary and inescapable reference to the principles that are referred to.

Indeed, reading the decision reveals a strong focus on the classical principles to which Artificial Intelligence must adapt.

From another point of view, reference is also made to the rights and guarantees that also protect the intimate sphere of the individual, erecting a barrier, a predictive shield against external intrusions. Thus, reference is made on this point to Articles 14 and 15 of the Constitution, protecting the inviolability of the home and of conversations and communications; at supranational level, reference is also made to Article 8 of the ECHR in the part that protects private life, to Article 7 of the Nice Charter, on the subject of safeguarding

¹³⁶ Other decisions to which we refer are also very interesting, TAR Lazio, III, 21 March, 2017, no. 3742 on the binding nature of the algorithm that manages the software relating to the interprovincial transfers of teaching staff; TAR Lazio, Rome, III bis, 30 October, 2017, no. 10805, on the subject of the automated procedure for the classification of scientific journals; Consiglio di Stato, VI, 19 January, 2018, no. 353 on the automation of the competition criterion; TAR Lazio, II quater, 28 June 2016, no. 7479 on the subject of state contributions to live performances; and, most recently, Council of State, VI, 23 January 2018, no. 456 and Council of State, VI, 5 December 2017, no. 5773 on macro-organisational acts that regulated mobility procedures in telematic form..

¹³⁷ Decision no. 8472 del 2019.

¹³⁸ On this issue, please refer to E. CARLONI, *I Principi della legalità algoritmica. Le decisioni automatizzate di fronte al giudice amministrativo*, in *Dir. Amm.*, 2020, 281 ss.; see also S. CRISTI, *Evoluzione tecnologica e trasparenza nei procedimenti "algoritmici"*, in *Diritto di internet*, 2019, 382 ss.

and protecting the right to privacy and, lastly, to the right of defence and the corollary of *nemo tenetur se detegere*.

Furthermore, reference is made to those guarantees that ensure dialectical and equal confrontation and verification of the reliability of the source of evidence¹³⁹; in this case, reference is also made to Article 111(2), (3) and (4) of the Constitution, to the right of defence itself and even to Article 6 ECHR.

As has already been reiterated by other voices, the Article 8 ECHR provision is a case of protecting fundamental rights when artificial intelligence comes into play. Indeed, it provides for quite advanced protection, however, presenting only one problem: the lack of a clear reservation of jurisdiction.

Other fundamental guarantees to be taken into account are the one provided for in Article 24(2) and the one provided as a fundamental guarantee in Article 13 of the Constitution.

Therefore, in an attempt to summarise, we see how the classical principles referred to in the Council of State's decision need to find a form that adapts to the new frontiers of artificial intelligence and some classical principles that (perhaps) should be rethought.

Clearly, this decision assumes considerable importance precisely in relation to the logical-legal considerations that follow from it and that, at the same time, identify the conditions of legitimacy to which the use of algorithms in the public administration's evaluation procedures must be subject. Indeed, with respect to the previous ruling by the Regional Administrative Court of Lazio¹⁴⁰, a further step forward is taken in this decision. In fact, the Council of State shows at the same time a greater openness towards the possibility of being able to use automated decisions, encouraging their use by virtue of the unquestionable advantages of automating the process¹⁴¹.

More in detail, the judgment then states that the use of an IT procedure that leads directly to the final decision should not be stigmatised, but instead encouraged. This is because it entails numerous advantages, such as, for example, the considerable reduction of procedural time for merely repetitive operations devoid of discretion, resulting then consistent declination of Article 97 of the Constitution. However, and this is strongly reiterated by the Council of State, the use of such procedures can not in any way be substantiated in an evasion of the principles that conform our system and that also regulate the conduct of administrative activity.

¹³⁹ As emphasised by M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, 2.

¹⁴⁰ Tar Lazio, sez. III bis, no. 9224, 10 September 2018.

¹⁴¹ These advantages are especially appreciable with regard to serial or standardised procedures involving the processing of large numbers of applications without any discretionary appreciation.

Indeed, the technical rule that governs each algorithm 'remains a general administrative rule constructed by man and not by machine'.

Lastly, it is specified that the algorithm must be considered an administrative computer act: this then necessarily implies compliance with a series of fundamental principles also identified at Community level, such as the principle of algorithmic transparency¹⁴².

Moreover, and this in itself represents a very important aspect that is reiterated in the decision: the algorithmic rule must not only be knowable in itself, but also reviewable by the administrative judge, who must be able to carry out a verification in terms of the logicity and correctness of the algorithm's decision outcomes¹⁴³.

Lastly, in fact, the Council of State ascertained the violation of the principles of impartiality, publicity and transparency, since "it is not possible to understand why the legitimate expectations of subjects placed in a specific position in the ranking list were disappointed"¹⁴⁴.

At the jurisprudential level, it is important to describe the Council of State's position in two different decisions. In the first decision in 2020¹⁴⁵, the Council of State was called upon to define the relevance and limits in the adoption of automated procedures in the administrative sphere, through the use of algorithms within an administrative procedure. The decision recognised the need to take into account three fundamental principles in the use of such computer tools. Firstly, the principle of knowability, according to which everyone has the right to know about the existence of automated decision-making processes concerning them. The second principle can be defined as the principle of non-exclusivity of the algorithmic decision. For instance, in the

¹⁴² This knowability must be guaranteed in all aspects: from its authors to the procedure used for its elaboration, to the decision-making mechanism, including the priorities assigned in the evaluation and decision-making process and the data selected as relevant. This is "in order to be able to verify that the outcomes of the robotized procedure comply with the prescriptions and purposes established by the law or by the administration itself upstream of that procedure, and so that the modalities and rules on the basis of which it was set up are clear - and consequently open to review".

¹⁴³ "The aforementioned requirement responds in fact to the inalienable need to be able to review how the power has been concretely exercised, in the final analysis posing itself as a direct declination of the right of defence of the citizen, who cannot be precluded from knowing the modalities (even if automated) by which a decision destined to affect his legal sphere has been concretely taken".

¹⁴⁴ In the same vein, the Council of State's subsequent ruling no. 8472/2019¹⁴⁷ on a similar case of interprovincial mobility of teachers intervened. The administrative judge, after reaffirming the principle that the "technical formula", i.e. the algorithm, must be translatable into the "legal rule" underlying it in such a way as to be legible and comprehensible, ruled that "the invoked confidentiality of the companies producing the computer mechanisms used cannot be of relevance, since by placing these tools at the service of the authoritative power, they obviously accept the relevant consequences in terms of the necessary transparency". The right of access to the source code of the algorithm, which is functional to the understanding of the functioning of the software, is thus reaffirmed, and a significant obstacle to the problem of the opacity of the algorithm is posed.

¹⁴⁵ Council of State, Sec VI, no. 881, 2020. On this decision, see A. VALSECCHI, *Algoritmo, discrezionalità amministrativa e discrezionalità del giudice*, in *Riv. Dir. Amm.*, 2020. This decision is in line with two previous decisions issued by the same section of the Council of State, Sec. VI no. 2270 and no. 8472, 2019. With respect to Judgment no. 2270, the Council of State takes significant steps towards a broader openness to the use of automated procedures even in cases of discretionary activities of the Public Administration.

case, in which an automated decision "*produces legal effects concerning or significantly affecting an individual*", the latter has the right to have that decision not based solely on that automated process¹⁴⁶. The third principle is algorithmic non-discrimination, according to which it is appropriate for the data controller to use appropriate mathematical or statistical procedures for profiling, putting in place appropriate technical and organisational measures to ensure, in particular, that factors leading to inaccuracies in the data are rectified and that the risk of error is reduced. In the present case, according to the Council of State, the algorithm does not appear to have been used in accordance with the mentioned principles; therefore, it cannot be considered legitimate.

The second decision took place in 2021 and is credited¹⁴⁷ with having addressed the sensitive technical question on the notion of an algorithm, which, in turn, has legal consequences. In this decision it is relevant to consider the introduction of the notion of "algorithmic processing" which is explicitly distinguished from the notion of artificial intelligence. In particular, the Court takes a position on the difference between the concepts of algorithm and Artificial intelligence and on the necessity of distinguishing the two notions. In fact, the Court establishes that it is possible to talk of "artificial intelligence" when the algorithm incorporates mechanisms of machine learning¹⁴⁸ and creates a system which does not limit itself merely to applying the software rules and the preset parameters (as the "traditional" algorithm does) but, also, constantly elaborates new criteria of inference between data and takes efficient decisions on the basis of such elaborations, according to a process of automatic learning¹⁴⁹.

There is an awareness of the fact that the use of modern technological tools actually entails a series of choices and assumptions that are anything but neutral: the adoption of predictive

¹⁴⁶ The introduction of algorithms and A.I. tools into the administrative process has been defined by scholars as the transition from "digital administration" to "algorithmic administration"; see D. U. GALETTA – J. G. CORVALÀN, *Intelligenza Artificiale per una Pubblica Amministrazione 4.0? Potenzialità rischi e sfide della rivoluzione tecnologica in atto*, in *Federalismi*, 2019, 6.

¹⁴⁷ Council of State, Sec III, no 7891, 2021.

¹⁴⁸ With regard to this last concept, in Italy there is no specific definition of machine learning systems. However, it can be deduced from some doctrinal writings that hold that machine learning and deep learning systems constitute the main learning methods of A.I. tools. The Italian translation of the term is 'automatic learning', which is understood as the ability of A.I. systems to learn, without any prior programming at all, therefore, without them being based on a precise model.

¹⁴⁹ One issue that has undoubtedly come to the fore is that the programmer of the machine learning algorithm is unable to predict how a given situation will change over time, and therefore cannot understand how the system has arrived at a given result. On these aspects, J. COPELAND, *Artificial Intelligence*, in S. Gutteplan (ed), *A companion to the Philosophy of Mind*, 1996, 124. The greatest risk associated with the outcome of the algorithm was also noted, which is that of passing on decisions and results that are imbued with bias. These are 'dysfunctionalities' that algorithms mostly learn from their programmer and the data fed into the system. See Vespignani, *L'algoritmo e l'oracolo* 103.

models and criteria on the basis of which data are collected, selected, systematised, ordered and grouped and coordinated, their interpretation and the consequent formulation of judgements are all, albeit reduced to automated operations, consequences of precise choices and values, conscious or unconscious; from this it follows that these tools are called upon to make a series of choices, which depend largely on the criteria and reference data used, regarding which the necessary transparency is demanded.

The 'multidisciplinary characterisation' of the algorithm is evident insofar as its elaboration requires not only legal expertise, but also technical, computer, statistical and administrative skills; once applied to law, it is then necessary that the 'technical formula', which in fact represents the algorithm, be accompanied by explanations that translate it into the underlying 'legal rule' and make it readable and comprehensible.

4.1.2 The Council of State's stance: the attempt to separate the concepts of algorithm and artificial intelligence

On closer inspection, an issue that has recently been animating the current debate concerns whether an A.I. system consisting of algorithms and data should not be considered as a whole, but on the contrary, whether it is necessary to distinguish between the concepts it includes.

This is because an attempt is being made, in the idea of classifying and distinguishing the different concepts more carefully, to distinguish A.I. systems that use machine learning methods. These, in particular, do not operate according to predetermined instructions, but adapt to new contexts and information, thus developing new behaviours not foreseen by the creator of the system.

It would seem that in part this position has also been taken by the recent decision of the Council of State¹⁵⁰, which stated that: "the common and general notion of algorithm brings to mind a finite sequence of instructions, well-defined and unambiguous, such that they can be executed mechanically and such as to produce a given result: nevertheless, if the notion is applied to technological systems, it is inescapably linked to the concept of automation, i.e. to systems of action and control suitable for reducing human intervention, the degree and frequency of which depend on the complexity and accuracy of the algorithm that the machine is called upon to process".

It is, however, a different concept when it comes to Artificial Intelligence where the algorithm contemplates machine learning mechanisms and creates a system that does not

¹⁵⁰ Sec. III, 25th November 2021 no. 7891.

merely apply software rules and preset parameters (as the traditional algorithm does), but at most constantly processes new criteria for inference between data and makes efficient decisions on the basis of such processing, according to a machine learning process.

However, it should be noted that, if one really wanted to draw a line between the 'algorithmic field' in the proper sense and the A.I. field, it would undoubtedly lead to some reflections, as has already been criticised and highlighted by the first commentators¹⁵¹: firstly, an A.I. system does not necessarily use machine learning methods (machine learning techniques, deep learning, etc.). To support this, in fact, the concept of A.I. includes systems that make inferences on the basis of knowledge representations provided by humans. Although this is a true starting assumption, today it is mainly machine learning systems that raise interest, expectations and concerns.

Secondly, then, also from a reading of the Council of State's decision, it emerges that even A.I. systems are based on algorithms for inference and learning. Indeed, in the case of systems based on machine learning, both the computer programme by means of which the system learns (the learning algorithm) and then the model by means of which the system responds to the input can be seen as algorithms, understood in a broader sense.

In conclusion, in approaching this particular field, it is considered useful that beyond the wobble in the relationship between algorithms and artificial intelligence, it would still seem preferable to broaden the concept of algorithm and then within it make distinctions with A.I. systems, based on the technologies that characterise them and the functions they perform.

4.2 *Predictive algorithms replace the judge: a look at the supranational landscape*

Already at this preliminary stage of the thesis work, it appears useful to emphasise how the application landscape and the first practical implications of the application of A.I. tools in justice are provided at supranational level¹⁵².

Indeed, there are some tools applied in the *online dispute resolution*¹⁵³ which are used to settle disputes in an entirely automatic manner or that, in parallel, have seen the introduction

¹⁵¹ Thus on this point, the first reflections in the margin by G. SARTOR, *L'intelligenza artificiale e il diritto*, 11. There have been cases in Europe where there has been an attempt to generate similar software, which, however, ended in failure. Reference is made, in particular, to the experiment conducted at the Courts of Appeal in Douai and Rennes, where software *Prédicitive* was tested for a short period of time: cfr. S. DUROX, *Des robots testés à la place des juges dans les cours d'appel de Rennes et Douai*, «www.leparisien.fr», <http://www.leparisien.fr/faits-divers/des-robots-testes-a-la-place-des-juges-dans-les-cours-d-appel-de-rennes-et-douai-30-10-2017-7362198.php>, October 2017.

¹⁵³ Consider that already a few years ago, 60 million disputes between E-bay traders were resolved through Online Dispute Resolution (ODR) software, which makes it possible to avoid recourse to the judge, and above all that some states, including Canada, but also numerous European countries such as Great Britain, the Netherlands and Latvia are progressively institutionalising these procedures, introducing more or less automated solutions for low-

of systems that replace (in all functions) the prosecutor in the decision on the assessment of the prosecution of certain types of crimes¹⁵⁴. The initial idea has, therefore, evolved to include algorithms with the aim of not only guiding and assisting the judge, but even replacing him¹⁵⁵. These are real 'robot-judges'¹⁵⁶ projects that it is deemed appropriate, at this preliminary stage, to at least describe in order to provide an idea of the current state of the art and current applications and then delimit the field of investigation of the paper.

4.2.1 *The Estonian case and the algorithm solving low disputes*

First of all, it should be noted that Estonia¹⁵⁷ has positioned itself as one of the most advanced countries in the field of technology and artificial intelligence¹⁵⁸. Indeed, it was the first state in Europe to introduce an A.I. tool within the judicial system. The idea behind these developments was to 'improve and implement' the justice system through the introduction of so-called virtual judges¹⁵⁹. Indeed, an algorithm was introduced that settles so-called low disputes, i.e. small claims, for as little as 7,000 euros.

value civil cases into the process. See, in this respect, the report by X. RONSIN- V. LAMPOS, *Studio approfondito sull'utilizzo dell'Intelligenza Artificiale nei sistemi giudiziari, segnatamente delle applicazioni dell'IA al trattamento delle decisioni e dati giudiziari*, in appendice alla *Carta Etica sull'utilizzo dell'Intelligenza Artificiale nei sistemi giudiziari*, available at the link: <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>.

¹⁵⁴ On this point, please refer to this Chapter, § 4.2.2.

¹⁵⁵ For instance, the scholar Richard Susskind, one of the world's leading experts on legal artificial intelligence, argues that it is possible that in 20 years' time, the judiciary may be completely replaced by sophisticated computer tools: *cfr.* R. SUSSKIND, *Online Courts and the Future of Justice*, Oxford, 2019. On this point, see anche M. R. COVELLI, *Dall'informatizzazione della giustizia alla «decisione robotica»? Il giudice del merito*, in A. Carleo, a cura di, *Decisione robotica*, Bononia, 2019, 125-137.

¹⁵⁶ A. TRAVERSI, *Intelligenza artificiale applicata alla giustizia*, cit., 3. On the possibility of applying these instruments in the decisional phase in order to overcome the scrutiny envisaged by Article 533(1) of the Code of Criminal Procedure, see G. CANZIO, *Il dubbio e la legge*, in *Diritto penale contemporaneo*, 2018, 1 ss.; M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, in *ivi*, 2019, 1 ss.; A. NATALE, *Introduzione. Una giustizia (im)prevedibile?*, in *Questione Giustizia*, no. 4, 2018, 1 ss.; in the same dossier, see also the contributions of C. COSTANZI, *La matematica del processo: oltre le colonne d'Ercole della giustizia penale*; also, C. Castelli-D. Piana, *Giustizia predittiva. La qualità della giustizia in due tempi*, in *Quest. Giust.*, 2018.

¹⁵⁷ Among the first to report this, E. NILER, *Can AI Be a Fair Judge in Court? Estonia Thinks So*, March 2019, in <https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/>.

¹⁵⁸ This is the idea of the Estonian government, which has already made considerable investments in digital transformation to the extent that it has become one of the most advanced countries in the creation of a 'digital society', with a transformation that began back in 1997 with heavy public investment aimed at creating IT solutions for e-government. In addition to online platforms for citizens in these more than twenty years, two generations of 'digital native' users and administrators have grown up, who see nothing revolutionary in creating systems to autonomously (and not trivially automatically) resolve the small legal disputes that clog up the courts in many countries.

¹⁵⁹ Indeed, Estonia as mentioned has a very long tradition of e-government. All public services are available online, one third of Estonians vote electronically. There is a national digital identity system, and it is linked to the digital signature that allows people to pay taxes, vote, have access to their health data, enter into contracts, and do online banking securely. In 2014, Estonia launched its e-Residency programme, an initiative aimed at making it easier to open a company or business in the country without having to live there. To date, some 50,000 people have gone through the process and take advantage of this almost complete digital flexibility as an easier way to

The idea behind it is that an '*automated robot judge*'¹⁶⁰ would be able, through the rapid processing of data, to manage paperwork and decision-making, thus making judicial services much faster and more powerful'. The decision made by the algorithmic system can still be appealed by humans and, in that case, the process would continue in the ordinary way.

The introduction of this tool was undoubtedly driven by the need to optimise and improve the speed of certain (*low value*) cases that were contributing to the slowdown of the Courts.

Therefore, one can see how the attitude shown by the Estonian state so far presents a situation in which this country is trying to exploit the new technologies not through other practices, but by trying to start a real project of transforming the delivery of justice for '*small claims*', exploiting an Artificial Intelligence system that has the task of comparing the data submitted by the parties.

It should be noted from the outset that the Estonian judicial system does not differ much from that of other European countries; however, it undoubtedly possesses the availability of an IT infrastructure such that bureaucratic-administrative and corporate life is so-called '*paperless*' and thus facilitates the introduction of a system not only capable of receiving the electronic documentation submitted by the individual litigants, but also of comparing it (and thus processing data) with regulations, filed deeds, regulations, smart contracts, in order to then issue a result.

The scheme and operation of the 'new technological process' are quite streamlined and simple indeed, the two parties in litigation submit the documentation in electronic format by uploading it to the site directly from their own office, with a number of indications as to what their respective legal claims and demands are. The Artificial Intelligence System analyses the documentation and the relevant regulations for the most relevant acts and, based on 'training' on cases with similarities (concerning already settled cases) and with the help of legally as well as digitally competent so-called 'coaches', issues a judgement, then agreeing with one of the litigants in the result and establishing damages.

enter the wider European market. Furthermore, the digitally fertilised earthquake of the public administration has given rise to a good number of digital tech companies in Estonia that have created a lot of wealth: one example is Skype, bought by Microsoft for USD 8.5 billion. But also, TransferWise for electronic payments and the ride-sharing service Taxify or Pipedrive for cloud sales.

¹⁶⁰ In particular, Ott Velsberg, Chief Data Officer of the Estonian government, was commissioned to design an artificial intelligence tool to deal with backlogged court cases in which the two parties upload documents and other relevant information and the artificial intelligence system issues a decision that can be appealed to a human judge.

4.2.2 *A look to the East: China and the new algorithmic prosecutor*

A technological future which maybe is not too far from the European reality is represented by the case of China, in particular the city of Shanghai, which has introduced a *prosecutor software* capable of analysing and processing data on, in particular, eight types of crimes¹⁶¹. This tool is intended to reduce the workload of the various Chinese district prosecutors' offices. However, the idea is to extend it to more types of crimes.

The machine was designed using 17,000 cases that occurred between 2015 and 2020 and would be able to identify the most frequent crimes committed in Shanghai. The researchers working on this tool believe that this type of software would have very low margins of error and would be able to present a charge and formulate it with 97% accuracy, based on the processing of data extracted from the case record¹⁶².

In particular, the three 'software skills' consist of: assessing the evidence, assessing the preconditions for arrest, and assessing the dangerousness of a suspect. It should be noted that the approach of this new technology is not entirely new to the Chinese system, as Chinese prosecutors had already been using software in the search for circumstantial evidence for some years¹⁶³. This new tool that is now being applied is capable of analysing and issuing a result in assessing the dangerousness of a suspect.

5 *The progressive approach to criminal justice in Italy*

At the moment, it should be noted that, there is no regulation in Italy in the various fields in which AI is applied. From the forecasts that can be made so far, following the majority approach on the subject, it is believed that an autonomous *ad hoc* regulation will be envisaged, using the categories of civil and criminal law for the individual sectors in which the introduction of A.I. tools will be envisaged. It is, therefore, envisaged that the future regulation will be differentiated according to the individual sector¹⁶⁴.

As already mentioned, nowadays justice systems, in general, have to deal with new actors and protagonists that move criminal law away from the typical '*purely human*' system, built on

¹⁶¹ In particular, its use is limited to the investigation of the following offences: credit card fraud, running a gambling operation, dangerous driving, intentional injury, obstructing official duties, theft, fraud and picking a fight and provoking trouble, an all-encompassing charge often used to stifle dissent.

¹⁶² In particular, the machine was built and tested by the Shanghai Pudong People's Procuratorate, the country's largest and busiest district attorney's office.

¹⁶³ This is the case, for example, with System 206.

¹⁶⁴ The topic is more about the use of artificial intelligence systems in the medical field, in the field of finance and insurance in the field of copyright law, in the field of self-driving cars and finally in the field of predictive justice or tools for crime detection in the criminal field.

man, and have to deal with new protagonists that seem to fascinate this field as well. In this paragraph, which serves as a premise for the deployment of the analysis in the following chapters, underlines the intention will be to attempt to provide a framework, for systematic reasons, of what are the possible uses, debated in part already by the doctrine of the Italian legal system, and then to assess the applicative aspects that we intend to analyse and, albeit in part, propose here¹⁶⁵.

The premise is that the *strong* evolution which has characterised the international landscape, affected by the unstoppable development of new technologies, has opened the way, even in the field of criminal law, to new problematic scenarios¹⁶⁶. As jurists, we witness as spectators the probable need for the already existing legal categories to adapt and readjust to new paradigms and needs arising from the development of new technologies¹⁶⁷. Undoubtedly, several issues emerge on which the debate is still raging in the doctrine, calling into play the main categories of criminal law: imputability, the structure of criminal liability, the concurrence of persons in the crime.

Criminal law is today undoubtedly called upon to pronounce on certain doubts still knotted and unresolved in the doctrine: on the level, for example, of the consequences of possible crimes committed by driverless cars, the need to satisfy the albeit legitimate claims of the victims can perhaps be traced back to levels other than criminal law, first and foremost that of compensation, which allows for imputative schemes that, in exceptional cases, disregard the principles - instead inalienable for criminal law - of liability for one's own deed and culpability¹⁶⁸.

¹⁶⁵ As has already been authoritatively stated in doctrine, the rapid advancement of artificial intelligence and its tools in the legal field has peculiar repercussions on the penal system. There are those among the Authors who have defined it as a real 'paradigm shift' in the face of the pressing on of certain tools that would seem to upset the foundations and structures of a system purely built on the human being and in which machines, at the moment, remain extraneous subjects. On the definition, we refer to the worrying denunciation of G. CANZIO, *Il dubbio e la legge*, in *Arch. Pen.*, 20 July 2018, 3 and also ID., *La motivazione della sentenza e la prova scientifica: "reasoning by probabilities"*, in G. Canzio – L. Luparia (eds), *Prova scientifica e processo penale*, Padua, 2018, 3.

¹⁶⁶ Emblematic is the definition given by F. STELLA, *Giustizia e modernità. La protezione dell'innocente e la tutela delle vittime*, Milan 2003, 292 ss., when it reiterates the need for criminal law to readjust and gear up to keep pace and not succumb to the impetuous emergence of new technologies that bring a 'shock of modernity'.

¹⁶⁷ In doctrine there are also those who have argued on the point that 'the idea that a machine, however 'intelligent', can - through the cold and algid consummation of an algorithm, by means of a 'robotic decision' - determine the fate of a person, whether it pertains only to the dimension of his assets, or even go so far as to regulate the level of affections and family relations, of compulsory health treatment, up to the decision as to the continuation or end of one's life, gives rise to concern and dismay. This, however, appears to be increasingly the future that awaits us and in respect of which the Jurist cannot - and must not - renounce his role, his presence, his work to know the phenomena, assess the concrete relevance and quality of the interests at stake and identify the appropriate disciplines and rules', E. GABRIELLI-U. RUFFOLO, *Dottrina e attualità giuridiche. Intelligenza artificiale e diritto*, in *Giur. It.*, 2019, 1657.

¹⁶⁸ A. CAPPELLINI, *Profili penalistici delle self-driving cars*, in *Riv. Trim – Dir. Pen. Cont.*, no. 2/2019, 341.

However, it is assumed that in the Italian legal system - at the moment - it would seem more plausible that on this matter criminal law would take a step backwards and that it would be up to, at most, other areas of law to regulate the matter, hypotheses and profiles that are more relevant.

Indeed, even before calling criminal law into play, it would seem appropriate to assess whether the categories and structure of imputation of civil and administrative liability can adapt to the new instances arising from the application of such instruments. Undoubtedly, it would seem that a very important role could be entrusted to administrative law¹⁶⁹, which could be called upon to deal with all the different chronological phases prior to the introduction of an A.I. instrument in a single sector. At the same time, it would seem plausible to hypothesise a form of insurance obligation aimed at regulating and protecting the subjects, right from the first phase of experimentation of the machine.

It is undoubtedly necessary to take note of a protection gap that also exists in other legal sectors¹⁷⁰. Undoubtedly, it is felt that the law dealing with artificial intelligence will have to be careful to promote the full development of its potential, while at the same time avoiding abuses and uses contrary to people's rights¹⁷¹.

In the current state, it seems necessary to further investigate the reasons behind this new right as well as its perimeter; a right that, as mentioned above, will have to be configured differently depending on the areas involved, the relevance of the decisions to be taken, and the overall balancing of opposing interests such as those of safety, efficacy and cost-effectiveness. In any case, it does not appear that the decision to maintain, in whatever sector one decides to introduce, a role for the human component, and consequently to avoid the development of full artificial autonomy, can be criticised as excessively conservative. It is also necessary to become aware that the need to regulate, and thus also to set boundaries, for technological applications of new scientific discoveries is a constant in the origin of bioethics as well as biolaw. In the case of artificial intelligence, it is a question of thinking of new legal principles and categories, or of modelling traditional ones, that can regulate in a balanced and proportionate manner a

¹⁶⁹ On the possible application scenarios of A.I. in public administration, see R. CAVALLO PERIN, *L'amministrazione pubblica con i big data: da Turin un dibattito sull'intelligenza artificiale*, in *Quaderni di dipartimento dell'Università di Turin*, 2021.

¹⁷⁰ On digitisation in the public sector, see a R. CAVALLO PERIN, *Ragionando come se la digitalizzazione fosse data*, in *Riv. Dir. Amm.*, no. 2, 2020. On this point also, I. M. DELGADO, *Automazione, intelligenza artificiale e pubblica amministrazione: vecchie categorie concettuali per nuovi problemi?*, in *Istituzioni del Federalismo*, no. 3, 2019.

¹⁷¹ C. CASONATO, *Potenzialità e sfide dell'Intelligenza artificiale*, in *BioLaw Journal*, No.1, 2019, 179.

phenomenon that is already present in everyday life and that within a few years is estimated to transform our very existence.

Undoubtedly, criminal law is at the centre of the debate when the main categories and institutes, related to the imputation of liability, causation, and the distribution of responsibility between the human being and the machine, come to the fore. The most relevant issues related to hypotheses in which the damage was caused by production defects of the system are, by way of example, all cases that fall outside the domain of the human controller. In such eventualities, criminal law could come up against insurmountable evidentiary difficulties involving certain founding institutions, such as the causal link and the principle of the personality of criminal liability, in its twofold declination of the prohibition of liability for the acts of others and the inalienable recognition of the judgement of reproachability. It follows that, unless one prefigures innovative flexibilisations and distortions of these institutions, a central role can only be assigned to the tort, which has always been open to the logic of objective imputation of damage¹⁷².

The pressing entry of artificial intelligence abruptly opens up a new scenario: criminal law will soon have to face new challenges imposed by the impetuous advance of technology¹⁷³.

Artificial intelligence and the use of algorithms aspire to penetrate to the roots of the system, touching the most diverse areas - from *policing*¹⁷⁴ to profiling to *sentencing*¹⁷⁵, in both *ante-delictum* and *post-delictum* perspectives -, and openly challenge the 'human factor' that informs the penal system: envisaging a 'legal-tech oracular system'¹⁷⁶ as an alternative. In fact, it aspires to improve the performance of the preventive and repressive system by operating at different levels¹⁷⁷, promising an exceptional improvement in effectiveness and efficiency, or even the definitive crowning of its objectives (the protection of legal goods); but, at the same time, it

¹⁷² C. PIERGALLINI, *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato*, 1756.

¹⁷³ For an overview of the scenario opened by the question 'machina delinquere potest??', see also, E. BASILE, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Dir. pen. uomo*, 2019, 27.

¹⁷⁴ C. PARODI-V. SELLAROLI, *Sistema penale e intelligenza artificiale*, in *Dir. pen. cont. — Riv. trim.*, 2019, 6, 47, 56; G. CONTISSA-G. LASAGNI-G. SARTOR, *Quando a decidere in materia penale sono (anche) algoritmi e IA: alla ricerca di un rimedio effettivo*, in *Riv. trim. diritto di internet*, no. 4/2019, 619.

¹⁷⁵ On this point see v. L. D'AGOSTINO, *Gli algoritmi predittivi per la commisurazione della pena. A proposito dell'esperienza statunitense nel c.d. evidence-based sentencing*, in *Riv. Trim. — Dir. Pen. cont.*, No. 2, 2019, 354 ss.

¹⁷⁶ V. MANES, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in (a cura di) U. Ruffolo, *Intelligenza artificiale. Il diritto, i diritti, l'etica*, Milan, 2020, 547 ss.

¹⁷⁷ U. PAGALLO – S. QUATTROCOLO, *The impact of A.I. on criminal law, and its twofold procedures*, in *Research Handbook on the Law of Artificial Intelligence*, W. Barfield e U. Pagallo (eds), Cheltenham-Northampton, 2018, 385.

envisages tensions with fundamental rights and genuine ethical challenges by undermining the fundamental principles of criminal law¹⁷⁸.

5.1 *The three application scenarios: investigative, evidentiary and decisional*

On closer inspection, the application scenarios which have most interested the doctrinal debate in recent years, branch out and intersect the investigative, evidentiary and decisional fields. Indeed, in the following paragraphs, the analysis will be limited and restricted to an attempt to provide a general overview of these major issues with which criminal scholars are confronted. The next few paragraphs will therefore serve as a brief description of the most emerging themes and, at the same time, serve as a premise for then assessing which direction has been taken in following a particular spectrum of application.

5.1.1 *Brief remarks on possible applications in the investigative field*

On closer inspection, the investigative field is one of the sectors where it is believed that artificial intelligence tools could represent a valuable resource. In particular, today we also speak of 'predictive policing'¹⁷⁹, meaning the set of all those activities that are directed to the study and elaboration of certain statistical methods which, applied to algorithms, are able to 'predict' subjects who might commit a crime, in order to prevent the commission of the same.

Taken as a whole, predictive policing contains a different set of analytical techniques, mainly quantitative, aimed at identifying likely targets for police intervention: the idea behind it is to solve crimes that have already taken place through statistical forecasts based on mathematical formulae (algorithms) that use and process huge amounts of data¹⁸⁰.

In this case, the prediction is based on a so-called 'actuarial reworking' of different types of data¹⁸¹; these include, in general, data concerning the individual and his or her routine activities,

¹⁷⁸ V. MANES, *L'oracolo algoritmico e la giustizia penale*, 548.

¹⁷⁹ "Such a distinction does not sound familiar outside the context of the English-speaking legal orders, not only because of the lack of a specific term, such as "policing", but also due to different legal conditions. Actually, the term 'policing' lacks of a precise translation in some of the continental languages, often being translated into the equivalent of 'surveillance'. Moreover, the term seems to represent a more active and independent role of the police, in the management of criminal files, being allowed to take initiatives that are not submitted to the control and permission of the judicial authority". For example, police in England and Wales leads the investigation until the final moment of prosecution, while in a certain number of continental legal order, the prosecutor is the master of the criminal investigation and the police almost responds to her guidelines. See on this definition S. QUATROCCOLO, *Artificial intelligence*, 37.

¹⁸⁰ It is necessary to see how this approach and type of technology to crime takes its cue, as a structure, from the software that is used to assess risks on social networks and those created in the health sector to predict from a statistical point of view various indicators that want to be examined.

¹⁸¹ According to a handbook developed in 2013 by the Rand Corporation for the National Institute of Justice (NIJ), and intended for law enforcement personnel at all levels, predictive methods can be divided into four broad categories: crime prediction: approaches used to predict locations and times when the risk of crime is

together with other data of an objective nature relating to reports of crimes previously committed, movements made, places frequented and the characteristics of the same; other data, on the other hand, concern elements 'external' to the individual, such as the time of year, weather conditions connected to the statistical occurrence of certain crimes¹⁸². Indeed, the intersection of these data, in some cases also combined with 'subjective' data, such as ethnic origin, level of schooling, economic conditions, semantic characteristics, can be traced back to subjects belonging to certain criminological categories¹⁸³.

In recent years, the use of such systems, which have been implemented especially at Italian police headquarters and also in other countries, has been gaining in popularity, allowing a better allocation of resources and an improvement in predictive policing. Through such systems, in fact, it has become possible to process an enormous amount of data, making it possible to acquire connections and information that were previously unattainable or that could be acquired only after a considerable delay¹⁸⁴.

5.1.1.1 The most popular algorithms as “crime-finding” tools: the case of Italian police headquarters

In Italy, as already mentioned in the previous paragraph, predictive algorithms have been created and used in several Police Headquarters.

It should be noted that *predictive policing software*¹⁸⁵ - whether assisted or not by A.I. systems¹⁸⁶ - can basically be divided into two categories "those that, drawing inspiration from the acquisitions of environmental criminology, identify the so-called 'hotspots', i.e. the places that constitute the possible *scenario* of the possible future commission of certain crimes - those that, drawing inspiration instead from the idea of crime linking, follow the criminal serialities

highest; offender prediction: software that identifies people at risk of committing a crime in the future offender identity prediction: techniques used to create profiles that match likely offenders with specific crimes that have occurred; crime victim prediction: used to identify groups or, in some cases, individuals who may become victims of crime. The report can be found at https://www.rand.org/content/dam/rand/pubs/research_reports/RR200/RR233/RAND_RR233.pdf.

¹⁸² For a complete framing of the subject of predictive policing, see W.L. PERRY-B. MCINNIS-C.C. PRICE, S.C. SMITH-J.S. HOLLYWOOD, *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*, Rand Corporation, 2013.

¹⁸³ See on this point, F. BASILE, *Intelligenza artificiale*, 10 s.

¹⁸⁴ *Ibidem*.

¹⁸⁵ Regarding I.A. activities, methods and tools, as well as their management, analysis criteria, risk forecasting and the development of police strategies for the use and allocation of human and financial resources, please refer to the OSCE document, Annual Police Experts Meeting: Artificial Intelligence and Law Enforcement: An Ally or an Adversary?, wien, 23-23 September.

¹⁸⁶ It is not always clear whether, and to what extent, the software we will discuss in the following pages is based on AI systems. This is also due to the fact that some of this software is privately owned and covered by industrial secrecy, so that details on how it works are not made public.

of certain subjects (identified or yet to be identified), to predict where and when they will commit the next crime. It must be said at once that, at least for now, both systems can only provide adequate predictions in relation to limited, specific categories of crimes (e.g. street crime, such as robbery and drug dealing), and not on a generalised basis for all crimes.

The two main software used in Italy, Key-crime and X-Law, are used within police headquarters to carry out so-called 'crime mapping' and identify where crimes may be committed. With regard to the Key crime algorithm, it analyses the characteristics of an offender and consequently, by analysing the data of his profile, statistically predicts where a crime of the same type may take place. The X-law software, on the other hand, concerns the detection of crimes such as theft, robbery and home invasion. In this software, the data collected concern the socio-environmental characteristics of the territory, the type of crimes committed daily, complaints filed, police reports, criminology studies, and much more, for which the machine returns a crime model with a prediction of the distribution of criminal activity over a given time span.

Therefore, the X-law software combines data pertaining to territory and type of crime. It works on a territorial basis and not purely with reference to the characteristics of the individual, with the specific aim of identifying so-called hotspots, i.e. areas where there is a risk of crime being committed. The programme relies on crime linking: it analyses thousands of data (from the where, how, when to the behaviour, clothes, means and weapons used by the robber) to establish which crimes have been committed by the same person or by the same group of subjects. It is in fact the crime linking to the criminal that establishes seriality in order to predict where the next actions will occur.

The best known artificial intelligence systems adopted by the Police Forces, which have reached a level of diffusion and frequency of adoption at a national level that allows them to be catalogued as operational standards, are: the "O.D.I.N.O" (Operational Device for Information, Networking and Observation) system used by the Carabinieri, the "MERCURIO" system of the State Police and the "X-Law" and S.A.R.I. (Automatic Image Recognition System) software adopted by the State Police. However, in addition to those currently being tested and/or developed, there are also a number of other I.A. systems adopted by the Police, often at their headquarters. Then there is the use of drones in certain circumstances: targeted checks, security of large events, search and rescue operations.

However, the most widely used software in Italy, Key crime and X-law, are only able to generate *crime mapping* on certain types of crime: in particular, theft, robberies and home invasions. The aim of the programmers of Key crime software is to be able to extend it to

crimes such as burglary and sexual assault. To date, data show that there has been a significant reduction in theft and robberies in places where these tools are used by police headquarters. In Milan, through the use of Key crime software, there has been a 58% reduction in robbery offences over the last two years. Undoubtedly, the improvement and reduction of robbery crimes have resulted in two main benefits: a not inconsiderable reduction in the commission of such crimes and, at the same time, a better allocation of resources followed by a better efficiency in the resources of the police and investigation apparatus.

5.1.1.2 *Brief reflections in the margin: problematic applicability profiles*

The applications of these tools which, little by little, have 'infiltrated' the police headquarters in Italy and Europe, has, from the very beginning, given rise to questions debated in doctrine, which have mainly concerned the problems that arise on the margins of the application of certain tools that process and process so much personal data of the subjects, making cross-references and connections. On closer inspection, the use of the same pertains to issues that cannot be read except through the first reflections on the margins of these uses. One cannot fail to notice that the use of such tools in itself presents a number of issues arising from the massive use of a number of personal data. Indeed, the use of such systems is limited to the use within the police headquarters of tools that are able to 'anticipate' and prepare the search for crime towards an improved allocation of resources.

In recent times, the use of A.I.-based software has enabled a quantum leap in predictive policing, since it is now possible to acquire and process an enormous amount of data, uncovering connections that were previously difficult for the human operator to detect¹⁸⁷.

The predictive policing systems briefly described above can undoubtedly bring great benefits in the prevention of at least some types of crime, but their use raises more than one perplexity¹⁸⁸. First of all, in fact, it should be noted that their use does not seem to have been regulated so far, in any country, at a regulatory level, so that the conditions and modalities of

¹⁸⁷ C. CATH-S. WACHTER-B. MITTELSTADT-M. TADDEO-L. FLORIDI, *Artificial Intelligence and the "Good Society": the US, EU, and UK approach*, in *Science and Eng. Ethics*, 2018, 505 ss.; L. BENNET MOSES, J. CHAN, *Algorithmic Prediction in Policing: Assumptions, Evaluation, and Accountability*, in *Policing and Society*, 2016, 1 ss.; G. MASTROBUONI, *Crime is Terribly Revealing: Information Technology and Police Productivity*, 2017, available online at this link; for a concise overview, in Italian, of A.I. systems aimed at predictive policing, see R. PELLICCIA, *Polizia predittiva: il futuro della prevenzione criminale?*, in *Cyberlaws*, 9 May 2019.

¹⁸⁸ The considerations contained in the remainder of the text elaborate on insights and reflections formulated by L. PASCULLI, *Genetics, Robotics and Crime Prevention*, in *Genetics, Robotics and Punishment*, Padua, December 2014, 192, and also R. PELLICCIA, *Polizia predittiva*, which refers, inter alia, to the research carried out on the subject, and the related concerns expressed by Human Rights Data Analysis Group (Hrdag), raccolte nel sito <https://hrdag.org/usa/>, alla voce "The Problem with Predictive Policing".

their use, as well as the evaluation and valorisation of their results end up being entrusted to practice alone, and thus to the initiative, sensitivity, and experience of police officers. Yet their use could entail serious friction at the very least with the protection of privacy (in view of the large amount of personal data collected), and with the prohibition of discrimination (to the extent that, for instance, they identify dangerousness factors linked to certain ethnic, or religious or social characteristics)¹⁸⁹. These are, then, systems that to some extent feed themselves with the data produced by their own use, with the risk of triggering vicious circles: If, for example, predictive software identifies a certain 'hot zone', police checks and patrols in that area will intensify, with the inevitable consequent increase in the rate of crimes detected by the police in that area, which will then become even more 'hot', while other areas, originally not included in the 'hot zones', and therefore not manned by the police, risk remaining, or becoming, for years free zones for the commission of crimes. Moreover, these systems call for crime prevention through active police intervention, though, therefore, a kind of 'militarisation' in the surveillance of certain areas or certain subjects, without, on the other hand, aiming at crime reduction through an action aimed, upstream, at the criminogenic factors (social, environmental, individual, economic, etc.).

Finally, one should not overlook the fact that most of these software packages are covered by patents filed by private companies, whose owners are, rightly, jealous of their industrial and commercial secrets, so that one cannot have a full understanding of the mechanisms of their operation, with obvious detriment to the need for transparency¹⁹⁰ and independent verification of the quality and reliability of the results they produce.

5.1.2 *In the field of evidence: brief remarks*

On closer inspection, when we talk about artificial intelligence and criminal law one of the central issues concerns the possibility of introducing special types of evidence, also called electronic evidence, which are nothing more than a subset of scientific and technological evidence¹⁹¹. The basic idea is to provide the adjudicating body with evidence that presents and

¹⁸⁹ On these aspects, A. BONFANTI, *Big data e polizia predittiva: riflessioni in tema di protezione del diritto alla privacy e dei dati personali*, in *MediaLaws* 24 October 2018; E. THOMAS, *Why Oakland Police Turned Down Predictive Policing*, in *Vice.com*, 28 dicembre 2016; J. KREMER, *The end of freedom in public places? Privacy problems arising from surveillance of the European public space*, 2017, in particular, refer to § 3.4.2, "Prediction", 269 ss.

¹⁹⁰ This concept is also referred to as accountability, i.e. its ability to account for how results were produced from certain inputs. On the point, KROLL J. A. –HUEY J. –BOROCAS S. –FELTEN E. W. –REIDENBERG J. R. –ROBINSON D. G. –YU H., *Accountable algorithms*, in *University of Pennsylvania Law Review*, vol. 165:633, 2017, 662; M. GIALUZ, *Quando la giustizia penale incontra l'Intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, 13 ss.

¹⁹¹ G. CANZIO, *Intelligenza artificiale, algoritmi e giustizia penale*, in *Sistema penale*, 8 January 2021.

possesses within it a technological security given by the mathematical result of algorithmic calculations leading to a certain result.

The objective pursued by the introduction of these new tools traces the choice behind the introduction and infiltration of artificial intelligence tools within the criminal justice system, which is represented, if nothing else, by the precise aim of implementing the quality of cognitive and decision-making performance of the judging body.

Well, it should be noted from the outset that the application of these tools in itself generates issues and problems that are difficult to overcome if one thinks of balancing the instances of these tools and counterbalancing them with the right of defence, which implies in itself, the exchange and dialectical confrontation, the possibility of refuting evidence, the right to contrary evidence and doubt in itself.

The main question revolves, precisely, around whether or not it is possible to guarantee the adversarial nature of the evidence, at the risk of not incurring a sort of 'dictatorship of technology' that bends the judge and the right of defence to the renunciation of fundamental pillars¹⁹². This needs to be taken note of is that science and technology are making a headlong incursion into the criminal justice system, raising and soliciting a great deal of fascination from legal practitioners.

The greatest doubts on which the doctrine has focused concern the acquisition of evidence: on the one hand, there is the idea of introducing these A.I. tools of evidence as a veritable filter at the end of which it is assessed whether or not to accept evidence; a second hypothesis, on the other hand, concerns deciding whether or not certain evidence produced by A.I. systems can be taken at trial.

In such a case, and precisely from this line of thinking, the first question would arise as to whether, with respect to the principles of procedural equality of arms at trial, the adversarial

¹⁹² On this point, reference is made to an interesting analysis by an American Court, in *Daubert v. Merrel Dow Pharmaceuticals, Inc.*, 509 US 579, 1993. In fact, although this case was not very recent, on this occasion the American Court had taken the opportunity to draw the boundaries and limits in the application of technological evidence. In fact, on the same occasion it states that the judge must examine the actual reliability of a theory or method and expert testimony for their admissibility as scientific evidence in the trial: the controllability by means of experiments; the falsifiability by means of negative disproof tests; the peer review of the scientific community of reference; the knowledge of the percentage of error of the results; and finally, the subordinate and auxiliary criterion of general acceptance by the expert community. It is interesting to note, as G. Canzio notes, how, following the same thrust, the Italian Court of Cassation with the *Cozzini* decision (Cass. Sec. IV, 17 September 2010, no. 43786), in substantially sharing the *Daubert* standard, has enriched its scope, with regard to the stage of the judge's assessment of scientific evidence, by adding the criteria of the independence and reliability of the expert, the breadth and rigour of the critical debate that accompanied the research, the aims and studies that support it, and the explanatory aptitude of the theoretical elaboration.

nature of evidence and the right to evidence and counter-evidence, there would be problems and questions as to their taking¹⁹³.

Being able to limit ourselves here to a summary analysis of the issue and with the sole aim of providing a brief overview of the possible implications of the A.I. in matters of evidence, it appears useful to recall that the provision under Article 189 of the Code of Criminal Procedure is a pivotal rule on the subject of evidence as it traces the original intention of the legislator and the will that science can help the search for truth, through a flexible trial system, capable therefore of adapting to the inclusion of scientific or technological evidence. The rule makes it possible to implement and 'open' the system also to new instruments by means of special criteria that could be encapsulated in the appreciation of the relevance, non-superfluosity and concrete suitability (fitness) of the evidence to ensure the ascertainment of the facts, without, however, prejudicing the moral freedom of persons. In this sense, this choice is left to the critical scrutiny of the judge. It is no coincidence that, with respect to the constitutional principles inherent in the trial stages, it is necessary to put the parties in a position to know the methodologies and instruments that will be applied in the ascertainment. Therefore, the judge, after hearing the parties on the manner in which the evidence is to be taken, provides for its admission, using the instrument of the order, and at the same time laying down the rules for the proper application of the methods and technical procedures for its acquisition. It can be seen from the outset that this provision functions as a "tighter filter" than the provision pursuant to Art. 190(1), which only negatively selects evidence that is expressly prohibited by law or that is superfluous or irrelevant¹⁹⁴. Moreover, this filter is assisted by a significant strengthening of the anticipated cross-examination "for the evidence", even before "on the evidence".

5.1.3 In the decision-making field: risk assessment tools

Furthermore, this is precisely the third strand or guideline along which this investigation is being conducted. If we look closely, parallel to the development of predictive algorithms, the idea has developed in various sectors not only of law, but also of finance, of management, of

¹⁹³ It is interesting to note the position that the authors had taken when rewriting the Code of Criminal Procedure. Indeed, one reads a passage in the Report to the Preliminary Draft of the new Code of Criminal Procedure of 1989, concerning the scope of Article 189 of the Code of Criminal Procedure: "*It seemed that such an articulated rule could avoid excessive restrictions for the purposes of ascertaining the truth, taking into account the continuous technological development that extends the frontiers of investigation, without endangering the defensive guarantees*".

¹⁹⁴ On this point, an authoritative commentary already cited, di G. CANZIO, *Intelligenza artificiale, algoritmi e giustizia penale*, in *Sistema penale*, 8 January 2021.

creating real indicators or markers of risk useful for making evaluations, by means of numerical results, aimed at predicting a given outcome in the analysis of various risks at stake.

Risk assessment in the criminal court is not just a simply a descriptive tool, “but an instrumental one from which serious consequences may follow, not least of which is preventive detention. Risk assessment plays a part in every decision to sentence an offender to indefinite detention”¹⁹⁵.

On closer inspection, it is precisely in this vein that these tools are inserted. In particular, these are peculiar applications of Artificial Intelligence that make use of algorithms that utilise a very large amount of data pertaining to the past of individuals and are able to identify recurrences characterised by a statistical base that is much broader than that of *human judgements*¹⁹⁶.

In recent years, there has been a veritable explosion in the use of algorithms in criminal justice in North America.

In this introductory part, therefore, it is considered useful to make only a brief mention of the background to the topic of interest in this paragraph. As will be seen, the subject will be dealt with in the following chapter in which risk assessment tools will be analysed, firstly analysed in terms of their intrinsic characteristics and functioning and, secondly, from a practical point of view in the application proposal to which this paper is addressed.

Here, therefore, we will limit ourselves to providing brief indications of the spread of these instruments. Following this line, therefore, it is useful to note that it is precisely the diffusion of these instruments in other legal systems¹⁹⁷ that has allowed the jurist to orient himself and outline their characteristics, their first uses and criticisms.

We will limit ourselves to saying that these instruments aimed at analysing and 'predicting risk' are used in various sectors in which it became necessary to assess, for application purposes, what their current and possible uses might be.

¹⁹⁵ See, Preventive Justice, 124.

¹⁹⁶ M. GIALUZ, *Quando la giustizia penale incontra l'Intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, 3.

¹⁹⁷ To give you an idea of the landscape that is albeit briefly described, see how Between 2012 and 2015, 20 laws in as many as 14 states 'created or regulated the use of risk assessments during the pretrial process'. For their part, a number of very important associations - including the American Bar Association, the National Association of Counties, the Conference of State Court Administrators, and the Conference of Chief Justices - have spoken out in favour of the use of such instruments in the pre-trial phase. For due doctrinal and jurisprudential references on this point, see the very recent article by A.Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, in *Duke Law Journal*, 2019, 1043 ss.; Cfr. A. Widgery, National Conference of State Legislatures, *Trends in Pretrial Release: State legislation*, March 2015; V. B.L. GARRETT – J. MONAHAN, *Judging Risk*, in *California Law Review*, *Forthcoming*, 10-11.

The necessary premises for the possible proposal of some of these tools in the field of criminal justice, would require, first of all, the verification of certain elements: first of all, the validity of the mathematical model whose application is proposed and, consequently, the selection and the data that would be used by the tool¹⁹⁸; in the same view, in a second moment, it would undoubtedly be necessary to ensure the transparency of the evaluation process carried out by the tool itself and the consequent possibility of challenging the reliability of the results (or output) of the algorithm.

Lastly, it would be necessary to compare and evaluate the actual use of such tools in light of the impact, limits and guarantees proposed by the Italian Constitution, in order to assess the limits and boundaries within which the application of risk assessment tools can be imagined. What must be avoided without a shadow of a doubt is that of creating a sort of 'scientificity' or 'penal determinism', which transmogrifies from criminal law of the fact to criminal law of the author on the basis of or on the influence of the analysis of data pertaining to subjective characteristics of the subjects and in which dangerousness is inferred from mere mathematical calculations.

6 *The spread of predictive justice*

On closer inspection, the spread of the concept of 'predictive justice' has infiltrated several areas of the criminal justice system. It is no coincidence that we use this term, which represents a large container in which various systems and applications can be found. When we speak of predictive justice, we are undoubtedly referring to the main concept of 'prediction' understood as the ability to 'see before' (in a literal sense).

The term '*predictive justice*' can actually be used to refer to a large container that includes several concepts and tools within it.

Indeed, in the field of machine learning, the term 'prediction' is used in broad terms to indicate 'any inference intended to expand the information available on a certain problem, inferences that may concern not only the future, but also the past and the present'¹⁹⁹. The use

¹⁹⁸ As he lucidly states M. GIALUZ, *Quando la giustizia penale incontra l'Intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati uniti ed Europa*, cfr. "la benzina che alimenta qualsiasi sistema basato sull'I.A. è costituita dai dati ed è fondamentale non solo la quantità, ma anche la qualità di questi. Ove il meccanismo lavori su dati imprecisi o inconferenti il rischio di produrre un output inattendibile (o peggio, discriminatorio) è elevatissimo".

¹⁹⁹ See G. SARTOR, *L'intelligenza artificiale e il diritto*, 131.

of computer systems to anticipate future events and behaviour can take place in different forms²⁰⁰, thanks also to different technologies.

Well, in an attempt to provide here a brief description of the uses and major debates on the subject regarding the concept of predictive justice, one could undoubtedly start from the concept that comes to the fore (or could come to the fore) for the public prosecutor, as a criterion of evaluation in relation to the useful exercise of criminal prosecution, in a perspective of recovery of the efficiency of the system²⁰¹, aimed at the timely, efficient and more effective handling of proceedings.

The idea of prediction also includes that particular application aspect that will be dealt with in the following paragraphs, which relates to all those cases in which the judging body is entrusted with the task of making prognostic evaluations concerning a given subject. In particular, with regard to the assessment of the "subjective dangerousness" or of the capacity to commit offences or of the possible risk of reoffending; in particular, there are several cases in which the judge is required to make a prognostic assessment, such as, for example, the identification of the dangerousness that is relevant in the precautionary measure pursuant to Article 274, lett. c) of the Code of Criminal Procedure, for the purpose of applying a security measure, pursuant to Article 202 of the Criminal Code, in relation to the phase of choice and commensuration of the penalty, in which the offender's capacity to commit offences must be taken into account pursuant to Article 133, paragraph 2 of the Criminal Code. Moreover, this assessment of dangerousness may also be relevant for the purposes of the recognition of the conditional suspension of the sentence, in the case of the granting of alternative measures to detention and in the cases of application of the prevention measures provided for by Legislative Decree no. 159 of 2011. These are applications that have very delicate profiles since every time the judge is asked to make a prognostic assessment, several risks come into play, mostly related to the distance from certainty and the proximity with the hypothetical concept, which, however, can have immediately negative effects for the subject, since, in most cases, these are measures that attack personal freedom.

Within the panorama of different artificial intelligence systems, here we will only mention the different models that exist today and that describe the interference or autonomy that a given

²⁰⁰ "Sistemi informatici per elaborazioni statistiche sono disponibili da tempo. Tali sistemi sono largamente utilizzati per la valutazione predittiva di casi individuali, in settori quali l'assicurazione e il credito. Per esempio, metodi statistici possono essere usati per determinare la probabilità che un individuo possa decedere in un certo arco di tempo, o possa non essere in grado di restituire il credito richiesto", *cf.* G. SARTOR, *L'intelligenza artificiale e i diritti*, 61 ss.

²⁰¹ On this point, please refer to a C. PARODI -V. SELLAROLI, *Sistema penale e intelligenza artificiale: molte speranze e qualche equivoco*, in *Riv. Trim – Dir. Pen. Cont.*, no. 6/2019, 56.

A.I. system follows. Indeed, there are known models of A.I. that follow machine learning: supervised, reinforced and non-supervised.

This reference will be useful since the most widespread predictive justice systems fall into the first direction. Indeed, these are systems in which the machine and the human being are the protagonists. On closer inspection, in supervised learning, the machine learns by and through supervision, i.e. through a certain phase of instruction or training in which it is given a large set of examples, each of which combines the description of a case with the correct answer to it. On this basis, the machine constructs a general model that is also applicable to new cases, albeit partially different from those in the training set²⁰².

What is noteworthy is that training a system does not necessarily require a human instructor to take on the task of providing examples of correct answers to the system. On the contrary, in several cases, the training set could be gathered 'more freely', e.g. from historical data concerning certain activities and data collected in the past.

Along the same lines, also in the case of hypotheses and models of predictive justice, the prediction of future judicial decisions, the examples consist of precedents recorded in case-law archives and each example associates the description of facts with a precedent with the decision that is taken by it.

6.1 *Predictability in jurisprudence as an incomparable value*

In the margin of the above reflections and descriptions, it is necessary to take note of a new channel on which scholars are now dwelling and which concerns the treatment of foreseeability in jurisprudence (a topic that has been much discussed to date and to which much value is attached)²⁰³. For a long time, predictability has been seen as a corollary of conformism and as an obstacle to that constant adaptation of jurisprudence to social reality, which is entirely physiological in a society subject to constant change such as the present one. This is a wrong assessment because what one wants to underline by *emphasising* predictability is the message of certainty and stability that comes out of it and the inevitable costs that changes in

²⁰² Thus, for a more complete description on the subject, see G. SARTOR, *L'intelligenza artificiale e il diritto*, 46.

²⁰³ On this point, which has already been dealt with in the previous paragraphs, it is considered useful to dwell on it; see a V. ZAGREBELSKY, *Dalla varietà della giurisprudenza alla unità della giurisprudenza*, in *Cass. pen.*, 1988, 1576; G. GORLA, *Precedente giudiziario*, in *Enc. Giur. Treccani*, vol. XXXVI, 1991; U. MATTEI, *Precedente giudiziario e stare decisis*, in *Dig. Disc. Priv. - Sez. civile*, vol. XIV, 1996; M. TARUFFO, *Precedente e giurisprudenza*, in *Riv. Trim. Dir. e proc. civ.*, 2007, 712; A. CADOPPI, *Il valore del precedente nel diritto penale*, Turin, 2007; A. CADOPPI, *Giudice Penale e giudice civile di fronte al precedente*, in *Indice penale*, 2014, 14 ss.; G. COSTANTINO, *La prevedibilità della decisione tra uguaglianza e appartenenza*, Report to the 11th Civilian Observer Assembly, 2016; L. SALVANESCHI, *Diritto giurisprudenziale e prevedibilità delle decisioni: ossimoro o binomio*, Report to the 11th Civilian Observer Assembly, 2016; F. VIGANÒ, *Il principio di prevedibilità della decisione giudiziale in materia penale*, in *Diritto penale contemporaneo*, 19th December 2016.

jurisprudence in various aspects produce. Not only that, but what one wants to avoid in the first place are unconscious contrasts. Contrasts that are far more widespread today than differences in case law. The areas on which the debate therefore focuses are therefore two: unconscious contrasts on the one hand and conscious differences in jurisprudential orientations on the other. In both cases, they must be addressed with the circularity of jurisprudence by distributing information and knowledge²⁰⁴.

This is a rule of enormous potential because it introduces a tool that stimulates knowledge and confrontation, without imposing hierarchical impositions, but at the same time giving responsibility to the various stakeholders. The presiding judge, who is obliged to hold meetings and deal with interpretations and differences in orientations, and the judges, who must be willing to get involved, discuss their orientations and assess their appropriateness. However, this is a virtuous process since, as I have found directly on several occasions, when faced with differences in interpretation, the judges with minority views have personally and spontaneously adjusted, considering it more costly for the section and the system to stick to their own interpretation, rather than to arrive at an office orientation and a related shared organisation, which is much stronger and more solid externally.

There are several, in many ways physiological hypotheses of fully conscious jurisprudential contrasts arising from different interpretative options. From the first point of view, information is the first fundamental datum that allows one to gain awareness and move onto the terrain of confrontation. Under the second point of view, on the other hand, confrontation is the fundamental ground for both verifying mutual theses and refining them. But even on this terrain, solutions are found that move in a direction of unity and predictability. In many cases, in the face of different theses, the interpretation of the Court of Cassation, or a minimally consolidated interpretation of the Court of Cassation, has been expected and solicited in order to adapt, overcoming seemingly irremediable conflicts.

Obviously, there are quite different situations: those of an already consolidated jurisprudence that only new events or profound reflection can cast doubt on and change, or those of orientations in the making and under construction especially in the face of regulatory changes, new rights and changes in society.

²⁰⁴ This is by no means a foregone conclusion. In particular, in large offices, also due to the monocratic nature of the vast majority of judgments, it is very difficult to know what is going on next door and even the manager has limited tools to know. Therefore, Article 47-quater of the Judicial Ordinance was introduced, which among the various tasks of the section president outlines that of taking care of the exchange of information on case law experiences within the section. In this regard, the Circular on the Tables, most recently the Circular on the Formation of the Tables of Organisation of Judicial Offices for the Three-Year Period 2017/2019 (Plenum Resolution, 25 January 2017) in Article 97 provides for the obligation of periodic meetings directed to this purpose.

The awareness that there must be is that predictability has two further formidable consequences: on the one hand, it contains demand and on the other it is a fundamental social message because every change or uncertainty about jurisprudence stimulates new, otherwise hopelessly seen questions.

Furthermore, it sends a social message because the predictability of timeframes and guidelines gives certainty to the community about justice and living law.

Well, since reference is made to a theme that will pervade the entire development of the thesis, it is considered most opportune and interesting to dwell on a concept that, although it recalls ideas and concepts that are very pragmatic and used from a practical point of view, nevertheless, at the same time, it calls to mind a concept that is capable of taking on various facets and assumes a real value²⁰⁵. Well, if one analyses the concept that will be able to pervade the entire elaboration, why, when speaking of Artificial Intelligence and the criminal justice system; in particular, is the idea of certainty and its pairing with uncertainty recalled, as well as the need for exactitude and justice that is ardently called for by society and the legal system. Indeed, it is precisely on this level, on the marshy terrain of human decision-making that cannot by its intrinsic nature represent something that goes beyond the boundaries of certainty by assuming and engendering the conviction of something not only certain but also 'just' that the concept of predictability is inserted, recalling in itself something that 'helps man', something that man needs in order to be able to decide. On closer inspection, the concept invokes considerations and reflections that call into question ethical, philosophical and legal issues. It should be said here that the added value of predictability in criminal justice is closely linked to the rationale and intrinsic nature of certain institutions.

As already mentioned, there are several cases in which the adjudicating body is confronted with decisions that are closely connected 'to the future' or to prognostic evaluations that either put it in a position to assess and 'foresee' certain possibilities in the future, with regard to the decision on the affliction of a particular penalty or security or preventive measure. In all these cases, it is foreseeability that takes center stage and assumes immeasurable value since, if adopted as a support tool for a judge, it could be useful and of necessary value in aiding and supporting a given decision.

²⁰⁵ One is reminded of the pre-Enlightenment dream of the 'calculemus' of G.W. Leibzin, who hypothesised that it would be possible to resolve all legal disputes through the construction of mathematical models of predictive justice; on this point, see G. W. LEIBNIZ, *Dissertatio de Arte combinatoria*, 666; in it. A. Artosi-B. Pieri-G. Sartor (a cura di), *Saggio di questioni filosofiche estratte dalla giurisprudenza e Dissertazione sui casi perplessi in diritto*, Turin, 2015, 200; indeed, he believed that "*quando orientur controversiae, non magis disputatione opus erit inter duos philosophos, quam inter duos Computatistas. Sufficiet enim calamos in manus sumere sedereque ad abacos, et sibi mutuo (accito si placet amico) dicere: calculemus*". In particular, the author's basic idea and aspiration was that it would one day be possible to resolve disputes, not through a trial, but through calculations; he imagined, in fact, a calculability of disputes through mathematical models, as a function of a predictability of the decision.

On closer inspection, it should be noted how predictability entails as a 'precondition' the knowability of case law. When attempting to frame the topic, it was immediate to mention a predictive justice system that, if functioning, could be able to bring about a qualitative leap for both economic agents and justice²⁰⁶. Today, there is much talk about predictive justice in various countries; it has also been described as a '*two-faced Janus*'²⁰⁷ because of the two contrasting aspects it possesses: on the one hand, the risks it would entail of reducing to a completely automated handling of what are also called *small claims*, and on the other hand, the possibilities of reducing and speeding up decisions on certain issues. Undoubtedly, as the experts also propose, it would be necessary - in assessing the possible risks and benefits - to envisage what the future uses might be, and then to arrive at a genuine governance and regulation of the matter. All this with a view to being able to enhance and exploit the possibilities of change in the predictability and transparency of judicial decisions, while maintaining the constitutional requirements and guarantees on the one hand of the judge's autonomy and, likewise, respecting the right of defence of those involved.

This is, on closer inspection, a debate that involves legal practitioners, magistrates, scholars and that intersects several aspects in itself.

At the same time, it is necessary to take note of the fact that a safe and intense dissemination of technology can, if however adequately controlled and regulated, help and assist policies to improve the efficiency of judicial governance: everything that is aimed at speeding up the times, less exorbitant costs and file modalities are examples that look favourably on the entry of technology.

Precisely on this basis, the question arises as to what the critical issues and the meaning to be attributed to an expression that is so widely used today, such as predictive justice²⁰⁸.

Predictive justice is therefore a very synthetic label that, like a large container, contains within itself a wide range of options which have in common the application of sophisticated technologies, both with analytical/inductive purposes (e.g. decision-making patterns, or behavioural patterns by analysing and processing data concerning cases and decisions that have

²⁰⁶ C. CASTELLI – D. PIANA, *Giustizia predittiva. La qualità della giustizia in due tempi*, in *Questione di Giustizia*, No. 4, 2018.

²⁰⁷ *Ibidem*.

²⁰⁸ According to the definition authoritatively provided by Antoine Garapon, 'predictive justice is still at the project stage', meaning that it is in an embryonic phase in which the possibility of applying technologies to certain legal, jurisprudential or judicial fields is beginning to be tested over time. We are not faced, as the author claims, with the application of software programmes that 'run' on platforms on which digital files are available and that have as their dominant function that of managing the passages, writing, revision, sharing, and validation of the file. These are algorithms that have as their field of application decision contents, texts of judgments, decrees, acts of the judge in general, jurisprudential databases, belonging to even very different systems and distant from each other in terms of cultural sensitivity. These algorithms are structured as functions whose topics are precisely these fields and whose results are cost 'probabilities', decision orientation, penalty ranges, compensation ranges, etc.

already taken place), and with prospective/predictive purposes²⁰⁹. In other words, therefore, it is not a matter of predicting the exact outcome of a judgement, but, on the contrary, of identifying the direction of the judge's reasoning. Since such reasoning never has the nature of a linear syllogism, but is composed of deductive inductive analogical steps, the prediction will be focal and not punctual.

Even so, however, the potential of the applications developed to date to realise what is synthetically defined as predictive justice is highly significant with respect to the impact, we can expect on three fundamental dimensions in the legitimisation of justice: the response to the demand for justice, timeliness, and consistency. At the comparative level, it is mainly the first two aspects that have received attention.

As we will see below, Italy, on the one hand, is turning its attention to consistency. paying attention instead to the Dutch and French experiences: the Netherlands qualifies as the first country among those in Europe with a civil law tradition to have launched a broad strategy to reform the justice system with a view to serving the citizen and economic society. In the beginning of 2002 and continuing to the present day, the path of transformation of the judicial system has seen the creation of a Council of Justice whose field of jurisdiction is the management and evaluation of the functioning of the system, the valorisation of the experiences of innovation and the exercise of a continuous (also possible due to the strongly consensual rationalities that characterise the Dutch public administration) regulatory function by the centre. While the countries of the South initiate technological innovations with a view to responding to functional and external pressure, the continental cases - and the Netherlands is in this sense emblematic - also introduce technological innovations through a rationality that we would qualify as internal cultural appropriateness. The cases of cyber-justice experimentation that deserve attention are essentially two, one of which is particularly important for comparison with Italy. The first concerns the on-line management of mediation cases, i.e. the extra-judicial settlement of civil disputes (family law). *Rechtwijzer* is the result of an elaboration by the University of Twente and Hiil (Hague Institute for the Internationalisation of the Law), a consultancy platform with an international presence in policies promoting the rule of law and based in The Hague. It is a telematic device that creates a two-way interaction between mediator user and legal assistant, in a totally digital and

²⁰⁹ Specifically, this means that propensities are identified and, on this basis, the probabilities are assessed with which the decision of the judge - in the case of judicial dispute resolution - or of the mediator - in the case of activation of ADR (Alternative Dispute Resolution) mechanisms - can be expected to converge on a point that we can define as focal.

dematerialised manner. It provides triage, counselling, mediation in the proper sense, and monitoring of the enforcement phase.

7 The database system in maximising predictive justice and the calculability of judicial decisions

As already mentioned in the previous paragraph, an attempt has been made to frame the value, under various facets, that predictability assumes, especially in certain delicate phases referred to the adjudicating body.

As already mentioned, the concept of prediction and prediction refers to and means 'moving' from known aspects of a case (e.g. an object, an event, a person) the so-called predictors (or also called independent variables or feature characteristics), to an unknown aspect of the same case, the target to be predicted (also called dependent variable or label).

Indeed, in recent years, there has been an increasing increase and wider use of computer systems that are used for predictive purposes. The use of machine learning techniques has become prominent, creating a synergy between data collection for the automatic creation of predictive models and machine learning-based applications.

Italy has a great experience and tradition in the field of case-law databases. The Electronic Documentation Centre (EDC) of the Supreme Court of Cassation was established in the late 1960s and early 1970s. It is very wide-ranging, containing archives, through which it is possible to carry out extensive and exhaustive research, not only jurisprudential, but also legislative and doctrinal, not to mention regulatory legislation, ministerial circulars, collective agreements, ordinances and municipal and other authorities' regulations.

Between the 1960s and 1970s, the Centre began to organise, in an automated manner (with the information retrieval system), the maxims of the Supreme Court of Cassation (in particular and legal documents), giving rise to the Italgiure Find system²¹⁰. Italgiure was one of the first database of the jurisprudence of legitimacy worldwide. Due to the completeness of the data (35 million documents) and search channels, the database of the Supreme Court of Cassation represents a fundamental channel as well as fulfilling the fundamental role of a public service of legal information technology. However, it never succeeded in having the desired application; indeed, the problems were many and proved insurmountable: the sheer voluntariness of the input and the difficulties of classification. So much so that the Merit Archive contained in

²¹⁰ The first public demonstration of the potential of the centre's computer was given on 21 March 1969, the new computer became operational on 1st October 1969 and the first connections with some judicial offices date back to 1973.

Italgire has not been fed since 2004-2005 and only at this time is it being re-discussed how to restructure it.

The creation of an archive of jurisprudence on the merits was re-proposed in a regulatory act by Article 7 (Database of jurisprudence on the merits) of the Ministerial Decree Organisational measures necessary for the functioning of the trial office (1 October 2015). This provision entrusts the Directorate-General for Information and Automated Systems of the same Ministry with the performance of all the activities necessary to ensure, as of 31 December 2016, the start-up of the Merit Jurisprudence Database and the usability of the data it contains on a national basis. The task given to the Dgsia (General Directorate for Automated Information Systems) is to carry out all the activities to enable the inclusion of classification metadata in the database and to facilitate the search of the jurisprudence contained therein, by enhancing the search channels²¹¹.

The discipline, in many ways appreciable, has encountered several obstacles. Firstly, the difficulty of being able to make usable and classify the jurisprudential archive already existing in the civil console. Secondly, creating the same archive in the criminal sector. These difficulties are not only technological, as the type of management entrusted with them would lead one to think, but also of content, precisely because they concern the classification of measures, which obviously requires clear parameters. In this respect, the allocation of the criteria for the selection of measures to the presidents of the Court of Appeal and the Court of First Instance is largely perplexing. If the intention, as the title of the provision states, is to create a database, presumably a national one, of case law on the merits, it is irrational that the criteria can differ from place to place. At most, one could decide on a core of common subjects and measures at national level and leave a subsidiary space for each Court of Appeal. But then the risk would not only be to discount different local views, but also to depend on different opinions, without promoting that national comparison of jurisprudence that is so fruitful and productive.

The topicality of the issue is also demonstrated by the very recent resolution of 31 October 2017 of the Csm: "*Guidelines aimed at identifying the modalities for reconstituting a database of merit case law*". The resolution moves in the direction of reopening the merit archive within

²¹¹ According to the provision, the criteria for selecting the measures to be included in the database are established annually by the president of the Court of Appeal or of the Court of First Instance who avail themselves for the implementation of those who carry out the training apprenticeship pursuant to Article 73, dl 21 June 2013, no. 69 or Article 37, paragraph 5, dl 6 July 2011, no. 98 or who are part of the trial office pursuant to Article 50, paragraph 1-bis, dl 24 June 2014, No. 90.

Italgiure Web with the aim of enhancing merit jurisprudence and stimulating a fruitful comparison between merit and legitimacy jurisprudence. As a criterion for the selection of the jurisprudence on the merits, it is proposed to include in the archive the measures concerning.

- decisions applying newly introduced provisions;
- decisions constituting the first application of innovative orientations of the Court of Cassation, all the more so if issued in unified sections;
- decisions constituting the first application of decisions of the Constitutional Court;
- decisions on matters not the subject of rulings of the Court of Cassation, provided that the archive does not already contain decisions of the same district on the same matter and of similar content;
- decisions constituting the expression of concrete solutions adopted by the courts of cognition on particularly important decisional themes.

The idea would be to include the measures in full and not in maximised form, both because of the difficulty of identifying homogeneous classification criteria and for reasons of simplification. The general supply of measures would take place through the national archives that are being set up by the Ministry or the district archives set up through the filing of sentences on consoles within the telematic process. This with a view to collaboration and synergy with the General Directorate for Automated Information Systems of the Ministry of Justice. The civil measures to be entered into the database would be selected locally by district structures and then forwarded to the centralised merit archive office. A new console functionality should be used for the forwarding by creating an application software on the assistant console that would allow the direct insertion of the selected measure from the local archive to the national archive of *Italgiure Web*²¹². The selection of measures would not be conveniently left to the individual magistrate drafters or section presidents (a method that had already proven to be unsuccessful in the past), but it would be provided that each office or section with the help of the trial office would be called, systematically, under the responsibility of the president to use the monthly sectional meetings also for the purpose of collecting the measures of interest, according to the selection criteria indicated above. The material identified would then be collected at district level by the District Innovation Office (which will be able to make use of a team operating at district level) for the implementation of the *Italgiure Web* merit archive. Again, at district level, a final check would then be made as to whether the selection criteria of

²¹² For the surveillance sector the Surveillance Office Information System provides the possibility to acquire all measures in 'pdf' format on a district basis. For the criminal field of merit the national archive of measures in non-digital format should be created with the help of the DGsia.

the measures identified for the archive are met and the material can be homogenised and refined. Another channel will be that of measures published in law journals. The search keys will have to be full text. The individual document before being entered into the archive will have to be processed by highly specialised subjects to be identified in private companies or research institutes connected to university institutions, adequately financed. The processing of the data shall ensure a very wide search field (by words/legal words, by parts of the judgement, by normative references, by jurisprudential references, by subject or sub-matter), as well as the insertion of links to other archives, the insertion of metadata, as well as the anonymisation of the documents.

Ambitious project that the Council also pledges to finance with its own funds already set aside. This project has some weak points, particularly at the local level. In fact, entrusting each district with the collection of measures comes up against several difficulties. First of all, in terms of resources, since the work is entrusted to the individual sections, which are supported by the trial office, and, as a centralisation, to the Rid (District IT Referents). This can lead to a spotty implementation lacking the necessary homogeneity and standardisation. It would probably have been valuable on the one hand to involve decentralised training, but even more so to reach a national agreement with the universities to have their input and support. But there is still opportunity and time to do this, given the open nature of the council proposal. Lastly, full synergy must be achieved between the database envisaged in the ministerial decree, on which DGsia is working, and the Council's project. It is clear that any duplication would be an intolerable waste.

The resolution also deals with one of the most delicate points that any database must address, namely the protection of privacy. It is well known that Articles 51 and 52 of the Consolidated Law on Privacy with regard to judicial measures provide that the person concerned has the right, by means of a specific and appropriate request before the definition of the degree of judgement, to have the clerk's office affix to the original of the judgement an annotation aimed at precluding the indication of his personal details and other identifying data of the measures intended for dissemination to a vast and indeterminate public. Anonymisation is also always necessary in matters where the publication of the judgment deals with sensitive data or would risk infringing the rights of the persons referred to therein. The First President of the Court of Cassation, taking up the "guidelines on the processing of personal data in the reproduction of judicial measures for the purposes of legal information" dictated by the Privacy Guarantor, in his document of 2 December 2010 prescribed, in the case of the reproduction of measures for the purposes of legal information, the adoption of measures aimed at obscuring the identification data contained therein when they relate to expressly defined subjects, such as

- by way of example - measures concerning minors, marital status, family, sex crimes or prostitution.

The experience of the Court of Cassation is particularly valuable by managing the only large database of national case law. It follows those judicial decisions, even if not final, do not need to be generally anonymised, which is indispensable only if they fall within the categories identified in the annexes to the aforementioned decree of the First President. Moreover, measures containing sensitive data will only be included in the database if they are of real legal and scientific interest. The anonymisation should be carried out directly by the magistrate who drafts them, even though the same resolution subsequently entrusts this task to private companies or research institutes that will be entrusted with the computer processing of the documents.

What emerges is an enormous caution that goes beyond what is required by law, not least because of the risk of sensitive data being published in violation of the law due to oversights or errors. So much so that it is envisaged that 'until practices are put in place that give maximum certainty regarding the protection of the above-mentioned requirements, that consultation of the archive of merit be reserved - at least in an initial phase - only to magistrates'. In this way, the value as a public service of legal information technology is lost, although what is proposed is not a renunciation, but a gradual path.

8 *The fragmentation of the judicial decision: new needs in a justice system*

The extended timeframes of the Italian judicial system and shortcomings in efficiency represent the constant complaint and grievance that affects the current debate; a reality of justice that is faced with often unreasonable times and timeless trials.

The paradox is realised when there is a constant heavy introjected by Italian judges and judicial offices that in recent years have accepted the challenge on time, constantly improving them and sometimes resulting in phenomena of blind productivism, indifferent to outcomes and quality. But on closer inspection, society's demand is more complex, namely to have speedy decisions, but at the same time quality, fair decisions. If it is easy to quantify and monitor time, it is not so easy for quality. In fact, the only tools we have to measure it objectively is the rate of resistance to further degrees of judgement as well as, probably, the level of social acceptance of decisions. Both are extremely insidious parameters, particularly the second. In fact, the rate of confirmation or resistance of a decision at subsequent levels of judgement is certainly significant, but it discounts the fact that appellate and cassation decisions are better only by convention, as well as by a series of parameters with which they are adopted (the panel, the greater experience of the judges, the relative distance from the fact). Inevitable

convention, also because a dispute must be put to rest, but still a convention, given that we often see disagreements and differing decisions arising even between the different sections of the Court of Cassation.

In any case, the examination of the outcome of proceedings both at first instance and on appeal is a very serious and interesting piece of information that should be known in every office and should stimulate and guide reflection in every section and seat precisely in order to improve quality.

Even more problematic is the analysis of the level of social acceptance of decisions, because it clashes and discounts multiple factors, including political and mass media factors. Not only that, but especially at a time like the present in which fake news and a triumph of populism over different skills and professionalism dominate, the risk would be to be conditioned by instincts, even before the opinions of the public. This, if anything, strongly emphasises the need for a communication policy of judicial offices that can explain and clarify their activities and orientations.

The lack of attention to quality is however evidenced by the simple fact that while it is very easy to find data and analyses on the time taken by justice (see in particular the fundamental ministerial censuses), there are in practice no public national data either on the outcome of proceedings, the rate of appeals, or the rate of resistance of measures.

Nevertheless, it seems appropriate to note that the alternation of judicial outcomes is becoming less and less socially accepted. The appearance is that the overturning of a first instance judgment on appeal is not the physiological result of a system of guarantees, but the negation of legal certainty, both because of the longer time periods and the debatability that each decision thus comes to have. The concept of judicial error has thus been extended beyond any measure of reasonableness. A miscarriage of justice includes any discrepancy from the final outcome, whether it relates to persons subject to investigations who were hit by precautionary measures and then acquitted, or first instance or appellate rulings that were denied at subsequent levels. It is not accepted that the system, being based on human beings and inevitably multiple interpretations, has provided for appeals precisely to minimise the risk of errors. Not only that, but that it is wrong to consider as an error the assessment that is made at a procedural stage (the precautionary stage) on the basis of elements that are different from those later on the merits. Or that the activity of the interpreter in a multi-source era such as the present is primarily an activity of reconstruction and coordination of applicable sources and regulations and as such is less and less a mechanical activity and more and more a path in which discretion and professionalism are enhanced.

Thus emerges the strong controversy on judicial errors that become an instrument of political attacks that are part of that never-abandoned stream of attacks against the judiciary,

its independence and its guidelines with statements that are disturbing and that inevitably bring discredit and mistrust on justice²¹³.

The first step should be to achieve more *awareness*, i.e. the constant monitoring of data and guidelines, their dissemination in judicial offices, and comparison. To give an example, avoiding unjust imprisonment due to expiry of terms should be very easy with the current computer systems, but the relevant ministerial programme has been under experimentation for several years and everyone has to make do with paper or handmade schedules. The second could be the focus on quality with the same emphasis that is rightly given to timing. Indeed, it should mean knowing the rate of confirmation of the decisions of the individual judge and the office, to discuss the guidelines, to verify what any anomalies depend on, such as an excessive rate of reforms or, as far as prosecutors' offices are concerned, an excessive rate of acquittals.

9 *Perspectives de iure condendo: towards predictive justice behind the scenes of the courtroom*

Case-law databases, monitoring and extraction of guidelines by subject matter and type of business make it possible to go outwards, building a bridge of communication and common interest with the community.

One can be able to know with reasonable certainty both the expected timeframe in a court or court for a case to be finalised and also what the subject-matter orientations of the office are. This ensures total transparency, but it also serves as a powerful signal to users and citizens and can contribute to the containment of demand: knowing in advance the timing and likelihood of an outcome leads to avoiding reckless litigation and seeking conciliations or settlements. This calls for making the timing subject-by-subject transparent and public, as well as drawing from the rulings issued more than maxims, principles of law and case law that can help those with similar problems. Obviously, this entails study and elaboration work that can only be conducted subject by subject and starting with a few subjects and then gradually expanding them. Not only that, but this can only be adequately achieved by including additional external resources in a notoriously 'poor' field such as justice. But this is a typical terrain on which the involvement of the universities could be valuable and possible, as they would thus see their resources enhanced and committed.

²¹³ It is reported that for unjust imprisonment the compensation paid since 1988 would amount to EUR 630 million for 24,000 people, but beyond the reliability of the figure, there is no question as to how to improve the quality and reduce such a worrying figure, at least on the surface, but this is used as a mallet against an evidently 'bad' judiciary without seeking reasons and remedies. What is not acceptable in these polemics is not to point out an existing problem, but to settle it in controversy, when what is needed is an overall intervention that seeks to make the system more reliable and of higher quality.

The other field that could be developed, on which there may well be misgivings, is that of deriving the probability of the outcome of a case in a certain court or court, in the light of precedents and guidelines. This terrain, which is very advanced in France, where it has begun to work with algorithms, charging for the analysis, may not be convincing due to the inevitable divergence between case and case and the great imprecision of such data, but it must be tested and verified by the test of facts.

It would therefore be necessary to focus on transparency, awareness and circularity of jurisprudence. By means of shared databases of jurisprudence, it is possible to achieve greater transparency of decisions, knowledge on the part of individual judges of the level of resistance and acceptance of their decisions, feedback from the president of the section and the office manager and more generally from the operators of the various existing jurisprudential theses, information for operators and citizens on the jurisprudential orientations of the office and the likelihood of acceptance of a request made in a certain matter or the possibility of an offence being recognised and the range of penalties imposed.

This would also be valuable in curbing demand and having an overall preventive effect as well as creating a relationship of transparency and active cooperation with the territory.

It is not a matter of controlling jurisprudence, but of gaining awareness of it and proactively helping everyone to improve their work. After all, the rule of that Article 47-*quater* ord. giud. does not exalt the conformism of decisions, but tends to avoid unconscious divergences, trying to overcome the inevitable dissimilarity that exists in the different jurisprudences through comparison and acquisition of the value of the predictability of decisions.

The model to be followed is biphasic: extremely elastic and open in the moment of elaboration and construction of an orientation following regulatory changes (as we know they are by no means episodic in our era), social changes and the creation and realisation of new rights. Founded on consolidation and *stare decisis* when there is a stable jurisprudence. Obviously, the stability of jurisprudence does not mean immutability, but requires in order to make possible changes deep reflection and adequate motivation, as well as an awareness of the costs involved in abandoning a certainty for the system.

The planning steps that can be proposed are gradual precisely to prevent the analysis of case law from becoming a moment of censure, and on the contrary may constitute enrichment for all.

The first move is to enable the individual judge to be aware of the outcome of appeals against his or her measures. This obviously concerns judgments and final orders, but also a very delicate field, on which it is good to extend this service, such as that of personal precautionary measures.

To ensure that the judge of the preliminary investigation can know not only the outcome of the appeal of his measure before the Court of Review or the Court of Cassation, but also the decision taken in the trial at first and second instance. The same applies to the review court. This element can help to verify which types and issues require more attention and improve our work overall.

The second step concerns the extraction of data, i.e. identifying types of matters and issues, verifying decisional outcomes, and following their progress in second instance and in cassation. In fact, identical extraction should be carried out with regard to both the rate of acceptance of requests for precautionary measure issues, the rate of their confirmation by the Court of Re-examination and the Court of Cassation, and the outcome of trials in which a person subject to investigation has been hit by a precautionary measure.

Such data, overcoming instrumental polemics on unfair detention and miscarriages of justice, would give us valuable indications on how we can, albeit within the inevitable different interpretations and perspectives, improve our activity.

The third step would be the creation of a reasoned ceiling, as such invaluable for guiding, consolidating, changing case law. A ceiling and a database that are the basis to allow and develop that jurisprudential comparison expressly provided for by Article 47-quater of the Judicial Order, built precisely to avoid unconscious contrasts and to favour the predictability of jurisprudence. The objective must be full awareness of the jurisprudential orientations of the different degrees and the timing of justice both internal and external. Internal to allow knowledge and comparison of the different orientations and verification of their resilience at subsequent levels of justice. External in order to be able to offer the community for each topic or branch of subject matter what the different orientations are and whether there is consolidated case law to rely on, as well as the foreseeable timeframes that venue by venue and subject by subject justice can offer. This allows real sector analyses, which help the sections to orientate their work and provide essential information to the outside world.

10 Concluding remarks: the appeal of predictive algorithms and a transforming criminal justice system

In the margins of these initial reflections and framing of the matter, one can see how it is necessary today to take note that criminal justice represents the last great ground for the conquest of the A.I. Years ago, it was perhaps truly improbable to imagine that the criminal justice system, which has always been impersonated and focused on man, could be infiltrated

by technological tools. Nonetheless, it seemed impossible from the outset that within the courtrooms one would start talking about good actors such as robot judges or algorithms²¹⁴.

Indeed, it is necessary to take note of how much the entry of new technologies capable of processing huge and complex amounts of data into this field simultaneously generates a double mood and opposite feelings. On the one hand, in fact, today's jurist is confronted with new subjects equipped with calculations and numbers that generate, at first glance, a sense of great bewilderment and, at the same time, fear of something not yet fully known or explored.

Secondly, from a more purely anthropological point of view for the criminal process, which witnesses a new confrontation between the human being, whose limits and boundaries are known, and the algorithm or other A.I. tools, whose full capabilities and boundaries of their applications are not yet known.

To conclude these premises, it should be noted from the outset that the relationships and interconnections that are generated between A.I. and law and criminal justice are manifold²¹⁵. Consider, for example, in addition to the aspects already discussed, the investigative, probative and evaluative field referred to the judge, the *connections*²¹⁶ with the activities of the public prosecutor²¹⁷; the investigative activity, in which it will be possible to distinguish the activation of new proceedings when there are already pending proceedings for which the I.A. is useful to consolidate the evidentiary framework and in which the acquisition and evaluation of the

²¹⁴ For an interesting further analysis on the possible uses and implications of the A.I. in the criminal justice system, see R. KOSTORIS, *Predizione decisoria, diversion processuale e archiviazione*, in *Sistema penale*, 23 July, 2021, 2.

²¹⁵ In this sense, one only has to think, for instance, of the possible management of 'pending court cases, in order to achieve a timely and efficient and 'uniform' prosecution. For example, on the organisational aspects of judicial offices, see, among others F. BASILE, *Intelligenza artificiale e diritto penale*, and ZIROLDI A., *Intelligenza artificiale e processo penale tra norme, prassi e prospettive*, in *Quest. Giust.*, 18 October 2019.

²¹⁶ In fact, the adoption of some A.I. tools to support prosecution offices has been discussed in recent years. Already today, there are some A.I. systems that are already in use that allow one to quickly obtain control of all the current pending cases in a given office and thus have the possibility of verifying any criticalities in real time at the same time. Thus, on the point S. M. GUARRIELLO, *Intelligenza artificiale ed attività del pubblico ministero*, in A. F. Uricchio - G. Riccio - U. Ruffolo (eds), *Intelligenza artificiale tra etica e diritti. Prime riflessioni a seguito del libro bianco dell'Unione europea*, Bari, 2021, 491. For example, a computer application currently in ministerial use called 'console penale' is very useful. It makes it possible to constantly check the progress of proceedings and to provide information on the total number of pending cases and the status of each individual case. Thus, it is possible to check, for instance, the expiry dates of preliminary investigations or those of the pre-trial detention phase, delays in the execution of proxies by the judicial police. In these cases, for instance, artificial intelligence would play an essential role in the scheduling of various activities, in meeting procedural deadlines, and in preventing situations prejudicial to citizens' rights of liberty from occurring. It is a module of the SICP (criminal cognition information system). It is, in particular, an IT tool used to manage the role and plan ordinary work activities for each criminal magistrate. Its operation is linked to the creation of a 'magistrate's desk' that provides the magistrate with immediate access to all the necessary information: role with the relevant deadlines (both investigation and precautionary deadlines and deadlines for filing sentences), diary, calendar of hearings and statistics. It allows one to customise one's role to match one's own personal patterns as closely as possible and immediately provides search, cataloguing and filtering functions.

²¹⁷ S. M. GUARRIELLO, *Intelligenza artificiale ed attività del pubblico ministero*, 488 ss.

elements acquired must take place in such a way as to make them usable in criminal proceedings for the purpose of identifying the perpetrator of the crime; in the possible role, already mentioned, of 'predictive justice' for the purpose of the possible determinations to be made regarding the exercise of criminal action or regarding the assessment of the dangerousness of a subject. In this case, 'automated decision systems' come into the limelight, even though they are not yet widely discussed or dealt with in the Italian legal system²¹⁸.

Well, as can be seen, there are several tools that can be used in the criminal justice system and that can implement the system in general²¹⁹.

In conclusion, in the face of a jagged landscape in which Artificial Intelligence could find space and margins of entry, even within the criminal justice system, we have chosen to, after showing (albeit briefly) a brief state of the art of the current applicative uses, in this paper we will deal with a particular type of A.I. tools applied in a peculiar phase of the proceedings.

It should be noted from the outset that, already at this early stage, it is undoubtedly apparent that the fascination exercised by algorithms in the field of criminal justice needs to be calmed and readjusted to the demands of society and the guarantees that must remain firm. Undoubtedly, in the course of criminal proceedings, the judge is called upon to make various

²¹⁸ On this point, please refer to F. BASILE, *Intelligenza artificiale*, cit., who states that 'Algorithms based on A.I. have also been used for some time now for decision-making purposes in the most diverse fields: these are the so-called automated decision systems, which are becoming increasingly widespread, both in the private and public spheres. Among the decisions that such algorithms can take are, of course, also decisions aimed at settling, or preventing, disputes and resolving controversies. Indeed, in this field, new technologies - thanks to the possibility of tapping into enormous amounts of data from sources such as case-law and legislative databases, collections of precedents, and the like - have already developed highly sophisticated devices that use game theory, positive outcome analysis, and negotiation strategies to resolve issues, thus employing a methodology that those involved perceive as objective and unbiased. These are alternative dispute resolution methods, often handled exclusively online, which, compared to traditional systems, result in reduced time and significant cost savings for both the parties involved and the decision-makers'.

²¹⁹Reference is made, for example, to another system for the management of individual proceedings, called TIAP-DOCUMENT, through which the entire case file is digitised. It is an application that was developed by the Ministry of Justice for the computerised management of the criminal file, with the possibility of integrating its contents at the various stages of the proceedings. Another system, called 'TOGA', has also recently been implemented. It is not a ministerial system and therefore the public prosecutor can use it privately. This system has recorded all the numerous criminal offences regulated by the Criminal Code and special legislation. In particular, it is a real database that breaks down each rule through algorithmic logic. In particular, it makes it possible to "calculate the type of penalty, accessory penalties, the admissibility of oblation, the admissibility of probation, the admissibility of prosecution, jurisdiction, the admissibility of plea bargaining, summons and notification, the admissibility of wiretapping and the maximum duration, compulsory expertise, the applicability, the overall terms and phase terms of precautionary measures, both custodial and non-custodial. It also calculates the ordinary and maximum statute of limitations for both the completed and attempted offence, also on the basis of recidivism; it calculates the deadlines for the investigation, notification, witness list and lawsuit, taking into account the holiday suspension. It follows the possible modification of the penalty based on regulatory updates and verifies the relationships with other offences: which ones are absorbed, absorb or concur with the offence in question as established by the Supreme Court'. This application is very useful as it allows precise knowledge for each offence, of the applicable institutions and of the coordination between the various rules, thus also ensuring a considerable saving of time and reducing possible errors. Moreover, it could also allow a more conscious and faster choice regarding the determinations to be made.

predictive judgments on the dangerousness of the defendant, which the legislature has sought to circumscribe and delineate through special *prognostic criteria*²²⁰.

It is believed that the jurist will nevertheless have to deal not only with the positive aspects of these technologies, but also with the initial limiting aspects that characterise them and that also lead the experts in the field to confront barriers that have not yet been overcome; In particular, while recognising the merits of machine learning algorithms in finding hidden associations between observations (more effectively than a human being would be able to do), there remains the limitation that through machine learning it is still not possible to find the fundamentals to undertake new actions that have never been tried, nor, at the same time, to be able to provide a causal explanation of observations that goes beyond the association between measured quantities. In fact, two crucial points in machine learning remain and remain: transparency and causality are two unresolved issues that go, however, beyond the purely technical aspects related to the accuracy of the so-called curve fitting, i.e. the goodness of the approximation of the formulae with respect to the collected observations (only set to increase). This is explained by the fact that, on the one hand, such 'powerful' algorithms are not able to explain in terms comprehensible (and even accessible) to man a given result or prediction issued; in this case (and this is the aspect on which the criticism is most focused today) this is only possible through a set of weights, and thus risk remaining impenetrable black boxes.

²²⁰ Please refer to the reflections of M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, in *Diritto penale contemporaneo*, 29th May 2019, 10 ss.

Chapter Two

Risk assessment tools: how they work and current applications

The need for an investigation independent of supranational examples

SUMMARY: 1. Foreword: a general overview on risk assessment tools. – 1.1 Brief historical reflection on origin in the US context. – 2. New challenges for the Criminal justice process? Brief historical evolution of risk assessment tools. – 2.1. The four generations of risk assessment tools over time. – 3. The definition and operation of risk assessment tools – 3.1. Criteria in the method of operation of risk assessment tools. – 3.2. Brief hints on risk factors in risk assessments. – 4. The combination of static and dynamic factors. – 4.1. A deeper look: the different types of risk in relation to factors. – 4.1.1. The clinical risk. – 4.1.2. The actuarial risk. – 4.1.3. The professional structured risk. – 5. A crucial step: risk identification. – 5.1. Static and dynamic risk factors. – 5.2. Dynamic risk factors. – 5.2.1. The criminogenic needs. – 5.2.2. The psychosocial needs. – 6. The I.N.U.S. conditions in criminal behaviour. – 6.1. A case in point: the Risk Need Responsivity model. – 7. The dual application front of risk assessment tools in criminal justice. – 7.1. The application paradigm of risk factors. – 7.1.1. In an *ante delictum* application perspective. – 8. Brief notes on *post delictum* application uses (*Segue*) – 8.1 An overseas perspective: current applications of risk assessment tools in sentencing and recidivism risk assessment. – 8.1.1. The US model: between evidence-based practice and the systems used. – 9. Other risk assessments used in the investigation phase. – 9.1. An example of risk assessment: the Public Safety Assessment (PSA). – 9.2. The PATTERN algorithm: the system to be taken as a model? – 10. Brief remarks: the anticipation of an initial operating proposal (*Segue*). – 10.1. The latest generation of risk assessment tools: why are they better than others in criminal risk assessment? – 10.2. Overcoming the intelligibility of decision-making mechanisms. – 11. When risk assessment approaches artificial intelligence. – 11.1. Two characteristics compared: accuracy and predictive significance. – 11.1.1. The risks related to implicit bias. – 12. First concluding remarks on the use of risk assessment in criminal justice: towards more 'fair treatment' justice?

1 Foreword: a general overview on risk assessment tools

After carrying out a description of the current panorama of artificial intelligence in justice, we assess the possibility of introducing artificial intelligence systems within the criminal sector and the reasons contemplating the advantages and risks associated with their introduction.

Now, as is well known, as already mentioned, the assessment entrusted to judicial bodies, concerns in many cases a prognostic assessment based on the issuance of a judgement concerning the future conduct of subjects. In real experience, such evaluations are entrusted to judges, who decide through their own and mere intuitive judgments that defer to the personal experience of each of them and to common sense. Indeed, there are decisions that imply an assessment that looks to the future in the strict sense, i.e. those concerning the application of security, prevention and even precautionary measures; however, the same decision entrusted to the judge on the treatment of penalties implies per se an assessment that looks not only to the present (the right penalty in relation to the act committed), but also and necessarily to the future (Article 27, paragraph 2 of the Constitution).

Indeed, especially this panorama forms the background to a new concept to which only a brief and rapid reference was made in the first chapter²²¹.

On closer inspection, we can already anticipate that the development of risk assessment tools has been discussed as a possibility in the application of the criminal justice system since the 1980s²²². From the outset, in fact, the importance that such tools could assume in a field such as criminal justice, in which many evaluations require a forward-looking assessment²²³.

It is no coincidence that the introduction of these tools has seen its first applications, in the field of criminal justice in the US legal system, both in a preventive key (in the study of tools that are aimed at the police force) and in a subsequent key, as a tool referred to the judge for the commensuration of punishment.

One must take note of the fact that machine decision-making is established and addressed as a factor that is increasing in our lives in several areas.

Indeed, the idea of calculation and a numerical result would seem to promise and ensure a kind of very high *accuracy* and *efficiency* gains, allowing for greater security and cost savings.

Indeed, the social impact of these technologies is of considerable importance since it would seem to address the citizen by ensuring a real reduction in human intervention that could improve the condition of the current state of justice and the impact on the individual's rights and opportunities.

It is undoubtedly necessary to take note, as already mentioned, that the boundaries of technological possibility and knowledge have moved much further in recent years. In the face of the idea of a 'just' sentence or a 'more exact' prognostic assessment of criminal dangerousness, it would also seem to move forward the wishes and ambitious goals of the human being²²⁴.

²²¹ Chapter I, § 3.1.

²²² Please refer to an interesting introduction on the subject by K. YEUG-M. LODGE, *Algorithmic regulation*, Oxford, 2019, 58 ss.

²²³ In US criminal doctrine, it is pointed out that '*risk assessment*' involves the use of actuarial and algorithmic systems to make predictions about the probability of future crimes. On this point, among others, M. STEVENSON, *Assessing Risk Assessment in Action*, Minnesota Law Review, vol. 103, No. 58, 2018, 314: «The term "risk assessment," however, usually refers to the use of formal, actuarial, and algorithmic methods of predicting the likelihood of future crime or misconduct».

²²⁴ "The progress, unimaginable until a few decades ago, that computer sciences have achieved in various fields is increasingly manifesting itself in the field of law as well, thanks to the development of software capable of delivering 'algorithmic decisions' to replace those of the human judge. Today's algorithms can even give greater certainty to the law and ensure the predictability of decisions, as is the case in the Chinese penal system, where the most advanced algorithmic systems generate a warning that the judicial decision adopted is not in line with the established jurisprudential orientation - thus limiting the judge's discretion - with the ultimate goal of ensuring the 'reasonableness' of the ruling and the uniformity of the pronouncements"; on this point, D. ZINGALES, *Risk assessment tools: una nuova sfida per la giustizia penale?*, in *DPC*, 12th December 2021.

Therefore, the analysis in the following chapter leads us to introduce the subject with an almost obligatory premise. Firstly, a historical introduction on the origin of risk assessment will be necessary, which must, necessarily start by pointing out how this is a theme that was born well before digital and technological development and was subsequently grafted onto it: the question of assessing the social dangerousness of a subject existed well before I.A. and algorithms.

As already mentioned, the difficult activity entrusted to the judge in having to 'judge' implies in itself the need to operate and place oneself in front of certain choices, requiring predictions. The activity of predicting is the most arduous activity required of a judging body that 'detaches' itself from the facts and must operate that inverse procedure that must allow it to construct a 'pyramid of information' that leads it to a certain decision.

While in many countries in Europe (England, Wales and Scotland for example) there is a constant and continuous development of risk assessment tools to support prevention and treatment²²⁵, Italy as a country still shows much reluctance and resistance. What we want in fact to consider in this chapter - keeping the pillars and guarantees of our legal system firm - is the possibility, only at a later stage - when the criminal trial is concluded and the subject has already been found guilty, in the phase of commensuration of the sentence, to introduce risk assessment practices; this opportunity could (perhaps) allow to structure a treatment and to plan an intervention that responds to the best evaluations made on the subject in the final idea of identifying the best treatment for the same for re-educational purposes.

Therefore, in implementing risk assessment it is, fundamental, firstly, to look at the experience of other countries and secondly to evaluate the actual benefits that risk assessment can bring in the area of criminal recidivism and violence; lastly it is important to consider how risk assessment could be more effective and to what extent the Italian criminal justice system can draw useful and significant results from an integration of risk assessment in the practice of assessing the offender and in the planning of prevention and treatment.

In conclusion, before delving into the analysis of these instruments, it is believed - as a perhaps over-optimistic perspective - that they can help professionals to better realise their

²²⁵ It should be noted that today, according to a recent study, thirty-nine US federal states have their own risk assessment tool. In England and Wales, OASys, is the offender assessment system routinely used in the National Offender Management Service (NOMS), within which different risk assessment tools are used, depending on the criminogenic needs of the offender being assessed. Other European jurisdictions do not seem to have institutionalised risk assessment instruments, but in the absence of more up-to-date data, it cannot be excluded that there are ongoing experiments at local level. On this point, S. QUATTROCOLO, *Sui rapporti tra pena, prevenzione del reato e prova nell'era dei modelli computazionali psico-criminologici*, in *Mimesisjournal*, No. 1, Vol. 22, 2021, 271; see also S. FAZEL, *Prediction of violent reoffending in prisoners and individuals on probation: a Dutch validation study*, 2019, 197.

interest in activating a targeted and *individualised* re-educational programme that can make the sentence respectful of the dignity and needs of the convicted person, without neglecting the interest of the law in 'protecting' society.

This type of 'tertiary prevention' that does not only look at the before but is directed towards the future is a field that one must enter and that is fundamental because it looks to the after²²⁶.

There is the underlying idea of re-evaluating and re-evaluating the third paragraph of Article 27 of the Constitution, in an attempt to give a new reading to that part of the provision that is in some cases forgotten. It is opportune to see how the punishment - understood in its most natural sense as that resigned to the legal decision - must not only be proportionate, pursuant to Article 133 of the Criminal Code, to the offender's capacity to commit offences, but must also be responsive having in mind the susceptibility of the convicted person in deriving a re-educative benefit from the same. And it is precisely in this sense that the individualisation of treatment follows on from punishment, taking on a scientific-clinical dimension as an achievement, assuming an ethical significance to the extent that it confronts the most modern acquisitions of the psychiatric-forensic, psycho-criminological and clinical sciences, integrating them into the programming of intervention, and eliminating, at the same time, any impressionistic, improvised, discretionary drift, which has often been present in this legal sphere: the executive dimension of criminal justice.

Indeed, the idea that we propose in this work is to analyse is the possibility of using risk assessment and of inserting it within the Italian legal context; this analysis will be carried out by attempting to embrace an international perspective that looks at evidence-based scientific research, clinical literature and psycho-criminological tools that can be used in Italy.

The use which is advocated is to try to activate and identify a re-educational programme that is targeted and individualised, so that it can make the sentence respectful of the dignity and needs of the sentenced subject. The idea is therefore to try, through the use and support of such tools, to plan *ad hoc* treatments adherent to the criminogenic needs of the individual. Underlying this is the conviction that not only an *ad hoc* treatment that is tailored to the different reality of each individual can better correspond and adapt to the guarantees proposed by today's society. Not only this, but also how the same can - at the level of so-called tertiary prevention - reduce recidivism.

²²⁶ G. ZARA, *Tra il probabile e il certo. La valutazione dei rischi di violenza e di recidiva criminale*, in *Diritto penale contemporaneo*, 20th May 2016.

In conclusion, this chapter does not have the ambition to be exhaustive on the subject. After all, as will be analysed in the following pages, the practice of risk assessment has a long history, and its methodological and applicative evolution is testified by the rich scientific productivity that characterises this field. Instead, the intent that moves and has moved the writing and the choice to undertake this direction is another: to try to set up and develop the practice of risk assessment also in Italy, in order to be able to intervene not only to treat the persistent and violent criminal individual, but also and above all to prevent his antisocial continuity over time. This would imply and entail enormous and relevant implications on a legal-social, scientific and preventive level.

It will be noted, and this applies as an incipit and fundamental premise to this second part of the work, that in reality no form of risk assessment and treatment intervention exists that is capable of satisfying all the criminogenic conditions that populate the reality of the courts. In fact, on a scientific level - *one size does not fit all*²²⁷; the inevitable consequence of this is that many violent realities will continue to remain uncontrollable and unpredictable, inexplicable, but in the face of this inevitable and (it must be said, necessary exclusionary statistic) there are (and perhaps are the majority) many other criminogenic realities that are instead explainable, assessable and preventable.

In conclusion, from a methodological point of view, an attempt will be made herein to account for some more technical notions (in particular on risk assessment tools), which are indispensable for a basic understanding of the phenomenon of designing predictive software, without, however, going into the details of such notions since, besides being difficult for a jurist to understand and access, they would risk losing the centrality of the subject of the paper.

The purpose of this analysis is to sound out the usefulness of such tools and the possible functions (at the level of mere support) that they might assume, highlighting the most evident application limits, always having regard to compatibility with the fundamental principles of the penal system.

1.1 Brief historical reflection on origin in the US context

The theme of predicting the risk of committing a crime has its roots in a particular socio-legal strand that developed in the United States in the early 1980s: the aforementioned Selective Incarceration movement²²⁸.

²²⁷ G. ZARA, *Tra il probabile e il certo*, 12.

²²⁸ *Supra*, Chapter I, § 1.1.1.

In short, according to these thinkers, the criminal justice system must 'identify, or select, a sub-category of individuals particularly predisposed to violence and recidivism and incapacitate them through long periods of incarceration'²²⁹.

In contrast, the issue of predictability in the US legal system goes back even further, as far as the 1920s: however, estimating a subject's dangerousness in a minimally accurate manner was extremely complex in those days²³⁰.

This problem also recurs for the advocates of Selective Disqualification, although this represents the first attempt to give an organised theoretical basis to risk assessment²³¹.

What is certainly noteworthy, given that the United States is the primary 'geographical' field of enquiry, is that it is in this country that the shift from a more 'actuarial' or evidence-based approach in sentencing has taken place, and that this shift has been possible due to several causes²³². Indeed, it has been recognised that it is precisely the use of actuarial risk assessment tools that would be able to reduce the very high levels of incarceration that have proven both discriminatory and costly over time. What certainly turned out to be, and in some ways was, the fertile ground on which these tools developed was the fact that we lived with a system full of sentences that were not fully defined in which the judge's 'intuition' possessed a fundamental weight²³³.

It should be pointed out that this theorisation²³⁴, which was unable to assert itself for the reasons set out above, provided an answer to an important question in the debate in North American doctrine and jurisprudence in the 1970s' concerning a more rehabilitation-focused justice: it increased the discretion of judges and aimed at individualisation of punitive treatment.

Unfortunately, this, although intended to ensure fairer justice, resulted in more discriminatory decisions towards minorities²³⁵.

²²⁹ D. KEHL – P. GUO – S. KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing, Responsive Communities Initiative, in Berkman Klein Center for Internet & Society*, 2017, 3.

²³⁰ *Ibidem*, 4.

²³¹ HARVARD LAW REVIEW ASSOCIATION, *Selective Incapacitation: Reducing Crime Through Predictions of Recidivism*, in *Harvard Law Review*, 1982, 96, 2, 511 ss.

²³² See K. HANNAH-MOFFAT, *Unpacking sentencing algorithms: Risk, racial accountability, and data harms*, in *Predictive Sentencing*, Struthers Montford, 2019, 179.

²³³ It was precisely this system, which was so focused, and which devolved to the judge any choice based more on subjective than on objective criteria, that left room for bias and led to strong inequalities in the execution phase, limiting accountability, transparency and consistency.

²³⁴ *Supra*, Chapter I, § 1.1.

²³⁵ D. KEHL – P. GUO – S. KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing, Responsive Communities Initiative, in Berkman Klein Center for Internet & Society*, 2017, 5-6.

Hence a return to the retributive paradigm that had dominated the scene until then and placed the focus more on the fact to be punished than on the perpetrator. This was, among other things, formalised at the regulatory level by the Sentencing Reform Act of 1984²³⁶, which prescribed federal sentencing guidelines with a specific structure and also set up a Sentencing Commission to supervise the application of it²³⁷.

But even this approach revealed its discriminatory character; so much so that it led to the era of mass incarceration, which to this day constitutes one of the most serious problems of the judicial system in the United States of America²³⁸.

This was mainly due to the shift to a not merely retributive, but also managerial conception, as well as to reasons of political *consensus*²³⁹.

After the failure of rehabilitation as the compass of punishment, it was considered more efficient to identify, classify and direct the so-called dangerous social groups, and this was now possible thanks to the development of new surveillance and control techniques: the theory of selective incapacitation was back, under the name of New Penology²⁴⁰.

Until the beginning of the new millennium, the prison imperative was the constant of US criminal policy, i.e. mass incarceration and very severe punishments, and contrary to expectations this did not reduce recidivism; on the contrary, the rate of re-incarceration soared; on the altar of efficiency and consensus a very high human cost was paid, and continues to be paid²⁴¹.

Studies of the high recidivism rates of offenders in the prison system led to a rethinking of this policy in the 2000s²⁴², and Congress passed a law allocating millions of dollars to fund programmes to facilitate the re-entry of offenders who had served their sentences in the community²⁴³.

However, not much was done to decrease the prison population and soon rehabilitation practices began to take hold again: 'What is old often becomes new again'²⁴⁴. In this context, so-called 'evidence - based' practices began to spread that is, the need to incorporate

²³⁶ H.R.5773 - *Sentencing Reform Act*, 6th April 1984.

²³⁷ It refers to the *U.S Sentencing Commission*.

²³⁸ D. KEHL – P. GUO – S. KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, 6.

²³⁹ *Ibidem*, 7.

²⁴⁰ MALCOLM M. FEELEY - JONATHAN SIMON, *The New Penology: Notes on the Emerging Strategy of Corrections and its Implications*, in *Criminology*, 1st January, Vol. 30, No. 4, 1992.

²⁴¹ C. KLINGELE, *The Promises and Perils of Evidence-Based Corrections*, in *91 Notre Dame L. Rev.*, 2016, 548.

²⁴² J. TRAVIS, *Reflections on the Reentry Movement*, *Cuny Academic Works*, Vol, 20, No. 2, 2007.

²⁴³ *Second Chance Act*, 2008.

²⁴⁴ C. KLINGELE, *The Promises and Perils of Evidence-Based Corrections*, in *91 Notre Dame L. Rev.*, 2016, 551.

quantitative and scientific methods into sentencing, particularly useful as a basis for decisions on possible future behaviour²⁴⁵.

As early as 1998, there had already been studies proposing criminal justice reform centred on the analysis of empirical data, concerning the police and medical sector²⁴⁶; now technological development provided the accuracy to make this possible.

Thus, in recent years, risk assessment tools have been recognised as the key instrument of criminal justice bail reform in the United States²⁴⁷. Indeed, such tools produce estimates that are believed to be more 'accurate' than those that judges can make, contributing not only to limiting unnecessary pretrial detentions and incarcerations of non-violent offenders, but also offering an apparent corrective to decisions that are potentially tainted by bias, whether conscious and acknowledged by judges or not.

2 *New challenges for the criminal justice process? Brief historical evolution of risk assessment tools*

On closer inspection, an area that has been particularly flourishing in exploiting the potential of algorithms is the one of risk assessment tools: in particular, when they are applied to the prognostic assessment of an individual defendant in a criminal trial.

A ben vedere, è necessaria una premessa metodologica per poter proseguire con l'analisi dell'argomento. Invero, il risk assessment è lo strumento per valutare la potenziale recidiva, ma in realtà non un termine non specifico poiché trova applicazione in varie tecniche e aree scientifiche diverse.

On closer inspection, a methodological premise is necessary in order to continue with the analysis of the topic. Indeed, risk assessment is the tool for assessing potential reoffending²⁴⁸, but it is a non-specific term since it is applied in various different techniques and scientific areas.

²⁴⁵ D. KEHL – P. GUO – S. KESSLER, *Algorithms in the Criminal Justice System*, 7.

²⁴⁶ LAWRENCE W. SHERMAN, Ideas in American Policing, in *Police Foundation*, July 1998; EVIDENCE BASED MEDICINE WORKING GROUP, *Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine*, JAMA, 2420–21, 1992.

²⁴⁷ By 2004, 28 states were already using risk assessment tools. The Supreme Courts of the states approved the use of these tools and in some cases even encouraged it, as in Kentucky, Ohio, Oklahoma, Pennsylvania and Washington. Today, they are used practically all over the country and in at least one county in every state for a total of over a thousand counties that use them. See, B. HARCOURT, *Against prediction: Profiling, policing and punishing in actuarial age*, Chicago, 2007.

²⁴⁸ The term 'evaluate' has been used as opposed to 'predict', so much so that in the literature the term 'prediction' has been supplanted by 'evaluation', so sharply on the point S. QUATTROCCOLO, *Sui rapporti tra pena, prevenzione del reato e prova nell'era dei modelli computazionali psico-criminologici*, 271.

However, risk assessment has also come to be confronted with criminal justice²⁴⁹, giving rise to various types of reactions.

The introductory assumption is that 'nothing predicts behaviour as well as (or better than) previous behaviour', such tools 'include a number of risk factors that may or may not be balanced, to provide a classification of dangerousness risk on different levels (low, medium, high), or a probabilistic score (a probability of recidivism occurring within a certain time frame), or both.

Here, as already mentioned, an attempt will be made to start with a brief *excursus* on the evolution of these instruments and then to understand how they have gradually become part of some jurisdictional systems.

It must be premised that the history of risk assessment is undoubtedly interesting for at least two reasons: the first is certainly linked to the discovery of the evaluative limits of unstructured clinical prediction; the other, on the other hand, is connected to the cautious, but increasingly interested, 'permeability' of the justice system, which would seem to be beginning to accept these scientific results in some areas of judicial practice²⁵⁰. Predicting who among the criminal population is most likely to re-offend has undoubtedly considerable implications from a social, clinical and legal perspective.

This work proposes to analyse risk assessment within the Italian legal context, trying not to limit itself geographically only to our legal system, but trying to maintain a supranational perspective in looking at evidence-based scientific research, clinical literature and possible tools that can be applied in Italy in order to help professionals and, above all the judging bodies, to identify a re-educative programme and a commensuration of the penalty aimed at identifying the treatment that can make the penalty respectful, first and foremost, of the dignity and needs of the convicted person and to operate, at the same time, that part of the preventive function that is implicitly put into practice with the choice of the penalty treatment. Indeed, risk assessment, like diagnosis, is able to indicate whether and how that individual is most likely to behave and react, and what can be done about it; it therefore has a preventive function and a function of guiding and guiding intervention.

²⁴⁹ Psycho-criminological risk assessment has become crucial in several areas of criminal decision-making, in the pre-trial phase, in sentencing, in relation to prison benefits and in the follow-up of psychiatric situations.

²⁵⁰ See J. MOAHAN, *The clinical prediction of violent behaviour*, in *Crime & Delinquency Issues: A Monograph Series, ADM 81-921*, 134, 1981. Reviewing the studies (available in the late 1970s) on the predictive accuracy of clinical assessments in predicting violent behaviour, he stated that the false-positive rate (in this case, criminals considered to be at risk of reoffending who did not relapse into violence) in the assessments of psychiatric and psychology professionals who were asked to identify which psychiatric-forensic patients were at risk of violence was excessively high: in fact, no more than one in three assessments were correct.

This is precisely why the prevention of violence and crime is one of the fundamental objectives of health and criminal justice. The possibility of making the assessments entrusted to the adjudicating body in the phase already of choice of treatment more accurate is a way to make them more 'personalised' and individualised and to be able, even more, to contribute meaningfully to the recovery of the persistent and violent individual and in this way to promote the well-being and future of these persons. To date, there is undoubtedly still a great deal of methodological inconsistency, evaluative imprecision and inefficiency in identifying those who could benefit more than others from targeted, specific and individualised interventions.

On closer inspection, much is already known about why a person commits a crime, but much remains to be known about the criminogenic mechanisms underlying the continuation of a criminal career and, above all, about possible remedies to mitigate this transition or relapse.

Risk assessment is placed precisely in a scientific-applicative space as a practice, in particular, aimed at the prevention of criminal relapse and the treatment of the persistent and recidivist offender. Pursuing this directive, the paradigm of criminal careers will be presented in the following pages: attention will also be focused on the concept of antisocial continuity in its various manifestations²⁵¹.

In an attempt to provide a definition, risk assessment can be identified as that so-called *'anticipatory and preparatory scientific practice of preventive, rehabilitative, supportive intervention, which is not disengaged from the 'treatability' of the antisocial and violent person but allows its planning. It is not an 'exact science', but it is a 'concrete science' able to help the expert to understand with whom, when, how, and on what to intervene, to reduce the risk of criminal relapse and violence.*

Well, already from these premises one can see the importance and fascination that risk assessment tools are capable of when approached in the criminal justice system.

2.1 *The four generations of risk assessment tools over time*

To date, there are about four hundred types of risk assessment tools in use; in fact, it is possible to mark the evolution in four temporal phases²⁵².

²⁵¹ “Understanding this distinction is crucial in that criminal behaviour cannot be examined outside the social context in which it occurs and cannot continue to be studied leaving aside the legal climate that defines and condemns it as such. These are unavoidable steps if we are to reduce the distance that exists between the 'legal world' and the 'scientific world'; two contexts that have hitherto managed to operate in parallel, often ignoring each other and rarely consulting each other, G. ZARA, *Tra il probabile e il certo*, 13.

²⁵² On the gradual evolution of risk assessment tools over time, see B.L. GARRETT - J. MONAHAN, *Judging Risk*, in *California Law Review* 2020, Vol. 108:451; C. D. STIMSON, *The First Step Act's Risk and Needs Assessment Program: A Work in Progress*, The Heritage Foundation, No. 265, 8th June 2020, 4 ss.

The first phase of risk assessment tools was the predominant one until the 1950s. It was characterised by an unstructured clinical judgement: this meant that all the assessment was left to the prison administration operators, to the various figures involved, such as psychologists, psychiatrists, who had the task of assessing, solely on the basis of their professional experience, the need to give coercive measures according to the risk they were confronted with. The peculiarity of this first phase, which then faltered over the following years, was that it was all based on the professional intuition of the individual. For this reason, the weak point of this first phase was precisely the unreliability of the data, which often credited factors that did not empirically correlate with criminal behaviour²⁵³.

While in the first phase the latter were based on informal procedures far removed from the logic of scores and statistical calculations operated on a collective basis and centred on interviews with the subject to be assessed, in the second generation of tools, dating back to the early 1970s, the assessment methodologies became actuarial in nature and were based on 'static' factors (such as age, gender and criminal record).

It was at the beginning of the 1970s, in fact, that the idea began to be talked about and to make headway that certain tools, related to the use and analysis of certain data, could assume great importance when applied in the judicial phase. The peculiarity of this second phase was the development of 'second-generation' actuarial instruments that were based on static factors, which could then be evaluated and added together to provide a recidivism risk index based on three different levels: low, medium and high.

It was noted early on that these instruments, based on a different approach to the previous ones, were capable of surpassing and supplanting them in risk assessment because they were based on more certain data linked to the present²⁵⁴. However, even this second generation had aspects that needed some adjustment. In fact, actuarial risk assessments also showed the first signs of weakness due to the fact that they were excessively 'static': they tended to show the individual as immutable, thus failing to readjust to the changes that he or she had and that would therefore affect his or her risk incidence, which had to be readjusted²⁵⁵.

²⁵³ 'In short, these instruments were scarcely transparent and structured and, therefore, could not be independently validated, leaving room even for arbitrary and biased procedures', thus on the point S. QUATTROCOLO, *Sui rapporti tra pena, prevenzione del reato e prova nell'era dei modelli computazionali psico-criminologici*, in *Mimesisjournal*, No. 1, Vol. 22, 2021.

²⁵⁴ However, it is worth emphasising that the risk factors used in actuarial instruments are more likely to predict, rather than explain, recidivism. And, in fact, these new instruments immediately proved to be more accurate in predictive terms and many states, in the USA and Canada, began to set up specific treatment protocols, in relation to the different scores of prisoners in the administration of actuarial risk assessments.

²⁵⁵ On this point, other criticisms that such instruments had aroused, and it is peculiar to see how the debate even at the beginnings of such introductions, focused on this aspect, some authors asserting that actuarial instruments

At an early stage, the categories used were considered 'first' or 'second level' (there were also those who spoke of real generations).

In this primary phase, the first and second generations of risk assessment tools became known and took root. In fact, while initially risk assessment was based on informal procedures far removed from the logic of scores and statistical calculations operated on a collective basis and centred on interviews with the subject to be assessed, in the second generation of tools, dating back to the early 1970s, the assessment methodologies became actuarial in nature and were based on 'static' factors (such as age, gender and criminal record).

Only between the end of the 1970s and the beginning of the 1980s, with the introduction of third-generation tools, a particular category of risk assessment tools was generated, characterised, within them, also by 'dynamic' risk factors (such as employment and educational status), which statistical risk assessment tools became 'RNA' tools (Risk and Needs Assessment tools), as they were aimed not only at identifying the risk-offence, but also at decreasing the risk of reoffending of the subject and at facilitating his social reintegration²⁵⁶.

In essence, this third generation, which was gradually becoming more widespread, was based on the view that both risk factors and criminogenic needs are relevant²⁵⁷. This view hits the nail on the head, as it allows risk factors and criminogenic needs to be integrated, and this relates and links past data with present and future data. This very approach, which was later termed 'structured professional judgement', sought to bridge the gap between the purely clinical and the purely actuarial approach²⁵⁸.

Finally, with the ground already paved and tested by the previous three generations, the fourth generation of risk assessment tools took hold, marking a peculiar phase. This type of approach had in fact, on the basis of past experience, focused the assessment on a much broader range of factors. Within them, in fact, risk factors, protective factors and reactive factors were taken into account²⁵⁹. Thus, the strength of this generation was precisely outcome-based: in

tend to be influenced by race and thus have strong discriminatory effects, because the actuarial method de-individualises risk assessment, classifying offenders according to unalterable group characteristics that potentially do not coincide with the individual. Thus, on this point, see K. HANNAH-MOFFAT, *Unpacking sentencing algorithms: Risk, racial accountability, and data harms*, 179.

²⁵⁶ S. QUATTROCCOLO, *Sui rapporti tra pena, prevenzione del reato e prova nell'era dei modelli computazionali psico-criminologici*, 172.

²⁵⁷ On this point, please refer to § 5.2.1.

²⁵⁸ Scholars welcomed these tools and the new approach that had emerged because, rooted in sound scientific knowledge, these methods also leave room for professional discretion and yield good practical results. In particular, their structure allows for adequate independent validation of results. See, Hart 1998: 121 ss.

²⁵⁹ G. ZARA, *Tra il probabile e il certo*, 12.

fact, it allowed for a more accurate analysis, in which non-criminogenic needs are incorporated into the overall assessment of the offender's functioning²⁶⁰.

The ones we know today and that are most widely used are the so-called fourth-generation risk assessment tools. In particular, these are so-called ANN models.

These models most in use today are more complex than the others because they have a dataset with a higher number of items to process than previous models; indeed, this feature would guarantee greater predictivity. Moreover, the fourth-generation risk assessment is also characterised by the maximisation of risk prevention and management through the preparation of specific treatment programmes for the subject being assessed²⁶¹.

The fifth generation of risk assessment tools²⁶², on the other hand, is represented by algorithmic machine learning systems that appear from the outset to be characterised by software programmed to perform certain tasks and that update their codes, step by step, each time they 'learn' from the results observed²⁶³. To date, the paucity of available information - due to the secrecy of the data on their operation²⁶⁴ - does not allow us to conclude in the sense of their more effective predictivity²⁶⁵.

In conclusion, contemporary risk assessment tools are systems that are based on so-called logistic regression and other statistical classification methods and are counted among the 'simple machine learning' tools²⁶⁶, which are in turn distinguished from 'real machine learning tools', represented by more complex and evolved algorithmic systems.

²⁶⁰ Although lacking a significant correlation with criminal behaviour, elements such as self-esteem, anxiety, victimisation tendencies (all non-criminogenic needs, in fact), can facilitate a successful treatment outcome, because criminal behaviour is the result of the complex interplay between cognitive, emotional personality, biological factors, environmental contingencies, within a framework of cost-benefit relationships. G. ZARA – D. FARRINGTON, *Criminal recidivism: explanation prediction and prevention*, 2016, 165.

²⁶¹ W. D. BURRELL, *Risk and Needs Assessment in Probation and Parole: The Persistent Gap Between Promise and Practice*, in *Handbook on Risk and Need Assessment: Theory and Practice*, Faye S. Taxman (eds), 2017, 26.

²⁶² A.S. NIEDERMAN (et oths), *The Institutional Life of Algorithmic Risk Assessment*, in *Berkeley Technology Law Journal*, Vol. 34, No. 3, 2019, 711-712.

²⁶³ A.B. CYPHERT, *Reprogramming Recidivism: The First Step Act and Algorithmic Prediction of Risk*, in *Seton Hall Law Review*, vol. 51, 2020, 339; ID., *Tinker-ing with Machine Learning: The Legality and Consequences of Online Surveillance of Students*, in *Nevada Law Journal*, Vol. 20, No. 2, 2020, 457 ss.; F. BASILE, *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, in F. Basile - M. Caterini - S. Romano (eds), *Il sistema penale ai confini delle hard sciences. Percorsi epistemologici tra neuro-scienze e intelligenza artificiale*, Pisa, 2021, 14.

²⁶⁴ According to the definitions of the US doctrine, only the latter systems fall under Artificial Intelligence, as they are endowed with the capacity for self-learning. While it is true that talking about AI necessarily means referring to algorithms, the opposite is not true, as not all algorithms are A.I. See J. VILLASENOR - V. FOGGO, *Artificial Intelligence, Due Process and Criminal Sentencing*, in *Michigan State Law Review*, Vol. 2020, No. 2, 2020, 296.

²⁶⁵ J.S. WORMITH, *Automated Offender Risk Assessment: The Next Generation or a Black Hole?*, in *American Society of Criminology*, Vol. 16, No.1, 2017, 281-288 ss.

²⁶⁶ A.S. NIEDERMAN (et oth), *The Institutional Life of Algorithmic Risk Assessment*, 713.

In the specific case of A.I. algorithms that 'learn from experience' and are furthermore susceptible to evolution even independently of human supervision, additional problems would arise over and above those already raised by the tools in use today. One of their specific characteristics is the secrecy surrounding the rules and functioning of these tools, which does not, however, make it possible to rule out the possibility that they are already being used in criminal risk assessment²⁶⁷. The reasons and motivations for which the Italian legislator would seem to be able to decide or envisage giving space and entry to algorithmic systems of risk assessment in criminal proceedings, should lie in the greater reliability²⁶⁸ and impartiality that this kind of tool would guarantee in the assessment of criminal dangerousness²⁶⁹, following the example of the USA, where statistical methods of risk assessment have been used for almost half a century²⁷⁰, on the assumption that they ensure assessments free from cognitive bias, which could characterise those of a judge in person. Today, however, the results produced by the use of such tools may lead us to state that such alluring promises do not seem to be fulfilled in practice in terms of the guarantees of greater impartiality that should result from them.

However, up to now the algorithms used to calculate the criminal dangerousness of a defendant would seem to reproduce the same biases that are already discriminatory in nature, since they are based on statistical calculations referring to a plurality of persons, who possess the same characteristics and are accumulated within the 'same risk classes'. They therefore also take into account elements that do not only concern and relate to the defendant, but often end up being influenced by ethnicity and socio-economic status²⁷¹.

Moreover, even when factors affecting only the defendant are taken into account (such as previous arrests or convictions, for example), this is not an 'individualised' assessment, but may reflect real 'trends' in criminal justice; this is precisely what happens in the US system, which appears from the outset to be characterised by a centralisation of police functions towards certain categories of subjects already deemed to be 'more at risk' of anti-social behaviour, which

²⁶⁷ J. VILLASENOR - V. FOGGO, *Artificial Intelligence, Due Process and Criminal Sentencing*, in Michigan State Law Review, Vol. 2020, No. 2, 2020, 301-302.

²⁶⁸ According to some studies, an algorithmic system for calculating recidivism would guarantee greater accuracy than that of a human being. A.M. HOLSINGER (et oths), *A Rejoinder to Dressel and Farid: New Study Finds Computer Algorithm is More Accurate than Humans at Predicting Arrest and as Good as a Group of 20 Lay Experts*, 2018, 50 ss.

²⁶⁹ For a critical analysis of this aspect, F. BASILE, *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, 21; M. AMISANO, *Profetica-mente: neuroscienze, intelligenza artificiale e previsione*, in F. Basile - M. Caterini - S. Romano (eds), *Il sistema penale ai confini delle hard sciences*, Pisa, 2021, 138 ss.

²⁷⁰ W. D. BURRELL, *Risk and Needs Assessment in Probation and Parole*, 23-24.

²⁷¹ This generates so-called 'algorithmic bias', which concerns situations in which an individual or group of individuals is unfairly favoured or discriminated against. See T. SOURDIN, *Judges, Technology and Artificial Intelligence: the artificial judge*, Cheltenham, 2021, 72.

results in significantly higher numbers of convictions for defendants belonging to these categories, thus leading to distorted outputs that do not correspond to reality²⁷².

Indeed, an initial criticism levelled at such tools is that risk assessment tools not only seem to reproduce the same errors as in the past²⁷³, since they deliver results 'contaminated' by factors even indirectly influenced by economic-social and sometimes even ethnic conditions²⁷⁴, but at the same time determine the risk of multiplying cases of biased outputs due to calculation or coding errors, even when they are developed precisely for the purpose of nullifying this risk, such as the most advanced risk assessment models²⁷⁵.

It is clear that in order to question why the subject of risk assessment tools is introduced, it is necessary to ask oneself some starting research questions, as already mentioned at the beginning of the paper. In particular, it is necessary to ask oneself the following questions that will serve as guidelines and directions during the course of the following chapters. One wonders why: should the legislature entrust the assessment of a defendant's dangerousness to an algorithmic system? What are the prospects of impartiality and accuracy that an output promises us in 'predicting' the risk of reoffending? And what limits can criminal law place on the technological evolution represented by risk assessment tools based on Artificial Intelligence? The results produced by the recourse to such tools in the US criminal system offer us the first answers to these questions, which will have to be taken into account if and when we find ourselves, also in Italy, confronted with these 'predictive systems', as innovative as they are dangerously in conflict with the constitutional rights protecting the accused.

²⁷² C.J.J. AVERY, *An Uneasy Dance with Data: Racial Bias in Criminal Law*, in *Southern California Law Review Postscript*, Vol. 93, No. 28, 2020, 32: «The fear is that, at best, algorithmic decisionmaking perpetuates historical bias; at worst, it exacerbates bias».

²⁷³ See C. O'NEIL, *Weapons of Math Destruction. How Big Data Increases Inequality and threatens Democracy*, Penguin Books Ltd., New York, 2016, 14, 40, 162. According to him algorithms would reproduce the mistakes of judges in criminal proceedings, since they are developed by a human being and since they encode rules and errors from past experience; S. BAROCAS - A.D. SELBST, *Big Data's Disparate Impact*, in *California Law Review*, Vol. 104, No. 3, 2016, 671 ss.

²⁷⁴ An example of this is the IOWA Risk Revised (IRR) tool, which also uses inputs regarding the defendant's employment, housing status and previous convictions, all factors historically influenced by social inequalities and racial discrimination. In the United States, one segment of the population significantly discriminated against by the use of this algorithm for calculating recidivism is African-Americans, who are statistically more prone to recidivism because there is a high number of convictions against individuals of this ethnic origin. However, it has to be taken into account that a conviction may also be based on the economic inability of the defendant to pay the court costs: Public Attorneys often push their clients to plea bargain, with the consequence that the percentage of convicted persons will inevitably be influenced by the economic resources available; as the authors point out B. KUTATELADZE, *Cumulative Disadvantage: Examining Racial and Ethnic Disparity in prosecution and sentencing*, in *Criminology: An Interdisciplinary Journal*, Vol. 52, No.3, 2014, 514 ss.

²⁷⁵ C.J.J. AVERY, *An Uneasy Dance with Data: Racial Bias in Criminal Law*, 32.

In order to answer these questions, an attempt will be made, at first, to provide a brief and sketchy overview of how these instruments work. Only in a second step will an attempt be made to understand how they might prove useful in the proposed analysis.

3 *The definition and operation of risk assessment tools*

Risk assessment tools find application in all those cases in which it is necessary to make forecasts under conditions of uncertainty at the outset and, in many cases, where the available analysis data are limited. It must be said that in the various fields in which they find application, which in certain cases require a prognostic approach and assessment, it is not possible to know all the risks one is faced with²⁷⁶.

Contemporary risk assessment tools are algorithmic systems based on logistic regression and other statistical classification methods and are counted among the 'simple machinelearning' tools²⁷⁷, as distinct from the 'real machine learning' tools, represented by more complex and advanced algorithmic systems²⁷⁸.

Risk assessment has gone through periods of fluctuating fortune: starting from an initial rejection by some who saw it as a form of colonisation of risk, or as an acceptable stigmatisation of those on the margins of a society that operates by social exclusion, or even as a practice useful only to exacerbate the climate of social concern and fear that arises instead from the illusory correlation between the seriousness of the crime and high criminal relapse, from the perhaps overestimated association between mental illness and violence, reducing it to a risky business. Indeed, even part of the legal world has been concerned about the use of scientific evidence that could cause the 'weakening' of the certainty of the law by directly affecting the judge's free conviction, thus perhaps introducing more technicality, to the detriment of a reduction of the human and subjective element. An attempt is therefore being made, through the use of such instruments, to arrive at a result more suited to optimal treatment, overcoming the limits arising from uncertainty.

The reasons for which the Italian legislator might decide to introduce algorithmic risk assessment systems in criminal proceedings should lie in the greater reliability²⁷⁹ and

²⁷⁶ M. DOUGLAS - A. WILDAVSKY, *Risk and Culture. An essay on the selection of technical and environmental dangers*, Berkeley, University of California Press, 1983.

²⁷⁷ A.S. NIEDERMAN (et oths), *The Institutional Life of Algorithmic Risk Assessment*, 713.

²⁷⁸ D. ZINGALES, *Risk assessment: una nuova sfida per la giustizia penale? La pericolosità criminale al vaglio algoritmico delle probabilità dell'esperienza statunitense*, in *DPU*, 9 December 2021.

²⁷⁹ According to some studies, an algorithmic system for calculating recidivism would guarantee greater accuracy than that of a human being. See A. M. HOLSINGER (et oths), *A Rejoinder to Dressel and Farid: New Study Finds*

impartiality that such tools would guarantee in the assessment of criminal dangerousness²⁸⁰; this optimistic reflection undoubtedly takes its cue from the example of the USA, where statistical risk assessment methods²⁸¹ have been used for almost half a century, on the assumption that they ensure assessments free from the cognitive bias that might instead characterise those of a judge in person. Yet, today, the results produced by the recourse to such tools may lead us to state that such alluring promises do not seem to be fulfilled in practice in terms of the guarantees of greater impartiality that should derive from them. Therefore, the application of these in the criminal justice landscape encapsulates the idea of preventing, as far as possible, the risk of violence.

One of the ways to proceed in this direction is to start with an evidence-based risk assessment. A quick search on the most important international scientific databases, using keywords such as 'violence', 'persistent crime', 'risk', 'recidivism', 'assessment', immediately provides a list of bibliographic references of tens of thousands of scientific publications on the topic. For those who are now approaching risk assessment, it is perhaps difficult to believe that interest in the subject was not always so.

Risk assessment, in general, has gone through fluctuating periods of fortune and application that have seen rejection by some who saw it as a form of *colonisation* of risk or an acceptable *stigmatisation*²⁸² or a practice useful only to exacerbate the climate of social concern and fear that arose from the illusory correlation between seriousness of crime and high criminal relapse²⁸³ and between mental illness and violence, reducing it to a *risky business*²⁸⁴. Critical responses also came from the legal world, which views with concern the use of scientific evidence that could 'dilute' the idea of certainty inherent in the law²⁸⁵, and from the clinical

Computer Algorithm is More Accurate than Humans at Predicting Arrest and as Good as a Group of 20 Lay Experts, Vol. 82, No. 2, September 2018, 50 ss.

²⁸⁰ For a critical analysis of this aspect, F. BASILE, *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, 21; M. AMISANO, *Profetica-mente: neuroscienze, intelligenza artificiale e previsione*, *ivi*, 138 ss.

²⁸¹ W. D. BURRELL, *Risk and Needs Assessment in Probation and Parole: The Persistent Gap Between Promise and Practice*, 23.

²⁸² G. UNDRILL, *The risks of risk assessment*, in *Advances in Psychiatric Treatment*, Vol. 13, 291–297, 2007.

²⁸³ A. HORSEFIELD, *Risk assessment: Who needs it?*, in *Probation Journal*, 50, 374–379, December 2003.

²⁸⁴ Thus, as it was defined by J. S. GLAZEBROOK, *Risky business: Predicting recidivism*, in *Psychiatry, Psychology and Law*, Vol. 17, Issue 1, 88–120, 2010.

²⁸⁵ The issue lies in the need to prevent 'scientific evidence' from turning into legal proof and the expert from replacing the judge, eliding his margins of free evidentiary assessment. Precisely with reference to this aspect, which is as interesting as it is complex, it is also important to emphasise the need for a precise distinction between rules of evaluation and rules of judgement (e.g. the beyond reasonable doubt rule). The following specialist literature on the subject can be consulted: A. BIANCHI – G. GULOTTA – G. SARTOR (eds), *Manuale di neuroscienze forensi*, Milan, 2009; G. CANZIO, *L'“oltre il ragionevole dubbio” come regola probatoria e di giudizio nel processo penale*, in *Rivista italiana di diritto e procedura penale*, 2004, 303; O. DOMINIONI, *La prova penale scientifica*, Milan, 2005; P. FERRUA, *Epistemologia scientifica ed epistemologia giudiziaria: Differenze, analogia,*

world, which sees in risk assessment a tendency to reduce intervention to pure technicality, depriving it of its subjective dimension.

On the contrary, on another front, welcoming reactions have emerged to the possibility of being able to recognise those specific criminogenic needs that increase the likelihood of committing new criminal acts. In the absence of a precise risk assessment, any accountability, treatment, re-education and rehabilitation response will be discretionary and inaccurate. Although scientific research immediately began to highlight this potential²⁸⁶, it was only a few decades ago that risk assessment began to carve out a space for itself in the psychiatric-forensic, psycho-criminological, and criminal justice systems.

In an attempt to provide a brief overview, albeit from a purely legal perspective, it will be necessary to frame the topic by providing a description more relevant to the more technical aspects.

In particular, it seems useful to provide a description of their technical operation.

As already mentioned, these are tools that are used as part of a risk analysis and management strategy.

Risk assessment tools are used, the assessment procedure of which itself includes a process involving analysis and forecasting in order to identify 'threats' or possible risks that might occur. At the same time, these tools also identify the limits and thus the perimeter within which some sort of 'permitted risk' is permitted, as well as the methodological and process corrections that can help (in most cases) companies prevent it.

In an attempt to provide a more specific definition, it should be noted from the outset that the term itself is generic and does not apply only to risk management. It only generally indicates an assessment to be conducted on a given aspect depending on the context. Risk assessment is generally aimed at identifying and analysing risks in order to understand what the priorities for action are and then, at a later stage, produce strategic actions to contain and mitigate them.

interrelazioni, in L. De Cataldo Neuburger (ed), *La prova scientifica nel processo penale*, Padova, 2007, 3-30; ID., *Il giusto processo* (3rd ed.), Bologna, 2012; ID., *La prova nel processo penale. Struttura e procedimento*, Vol. I, Turin, 2015; U. FORNARI, *Al di là di ogni ragionevole dubbio. Ovvero sulla cosiddetta prova scientifica nelle discipline psicoforensi*. Turin, 2012; G. Gulotta, *La responsabilità penale nell'era delle neuroscienze*, in A. Bianchi - G. Gulotta-G. Sartori (eds), *Manuale di neuroscienze forensi*, 3-14; B. LAVARINI, *Neuroscienze e processo penale. Relazione ad un incontro di studio seminario specialistico presso l'Ordine Avvocati di Pinerolo*, October 2012; I. SINGH – W. P. SINNOT-ARMSTRONG - J. SAVULESCU (eds), *Bioprediction, Biomarkers, and Bad Behavior. Scientific, Legal, and Ethical Challenges*. Oxford, 2013; G. UBERTIS, *La prova scientifica e la nottola di Minerva*, in L. De Cataldo Neuburger (ed.), *La prova scientifica nel processo penale*, 83-91, Padua, 2007; G. UBERTIS, *Il giudice, la scienza e la prova*, in *Cass. Pen.*, 4111-4119, 2011; G. UBERTIS, *Profili di epistemologia giudiziaria*, Milan, 2015, 176- 177; G. UBERTIS, *Argomenti di procedura penale*, Vol. IV, Milan, 2016, 94-95, 244 ss.

²⁸⁶ D. L. SHAPIRO – A. M. NOE, *Risk Assessment. Origins, Evolution, and Implications for Practice*, Berlin, 2015.

The functioning and application of these tools implies, in itself, a subdivision into several stages:

1. Identifying risks.
2. Decide who may be harmed and how.
3. Assess the risks and define mitigation actions.
4. Record the results and implement the risk mitigation plan.
5. Periodically review the assessment and update it if necessary.

Even if we stop at a merely cursory description of how they work, what should certainly be noted is that they are tools that, if equipped with self-learning mechanisms, fall into the *macro* category of artificial intelligence tools. They would, therefore, be able to work autonomously in the reprocessing of data (in the final phase of the various stages). It should be borne in mind from the outset that risk assessment does not have univocally codified procedures; on the contrary, the phases are flexible and variable, but it is nevertheless considered good practice to start by studying the context in which one operates and try to identify the main sources of 'risk or threat'.

3.1 *Criteria in the method of operation of risk assessment tools*

Moving towards the heart of this chapter, an attempt will be made to provide some more precise indications on the functioning of risk assessment tools.

In the study of such instruments, as anticipated at the beginning of the chapter, an attempt was made to look at them as possible means of help in improving - in the Italian justice system - the treatment choice of the individual²⁸⁷, which implies within it a prognostic assessment that looks to the future and analyses the subject's capacity to commit offences.

Well, here it is considered necessary to *focus* on the method used by such instruments since it is believed that the same and 'measurement are two important and irreplaceable steps of any scientific investigation, without which any subsequent exploration would be impossible'.

The application model, as already mentioned, is the Risk Need Responsivity (RNR), that will be described in the following paragraphs²⁸⁸. The basic idea is precisely that of not annihilating or simply 'detecting' a certain criminal behaviour or level of dangerousness but is

²⁸⁷ *Segue* in more details in Chapter 4.

²⁸⁸ As already anticipated, the model that is proposed in terms of deciding the sanctioning treatment is that of the interested, motivated, renewed participation of the antisocial individual who is no longer understood as "a passive recipient of the intervention", but instead "a subject who actively cooperates in the treatment programme, contributing to making it individualised and specific."

that of identifying it in an 'active sense' with a view to achieving the twofold preventive aim of neutralising but also looking to the future.

Starting from these premises, however, it is undoubtedly not possible to abandon the awareness that every form of assessment in itself implies a dose of subjectivity (referred in all cases to the judging body) and that - perhaps precisely for this reason - it is fundamental and necessary to start establishing a method, parameters and criteria that are able to guide the expert. For this reason, one turns and must be oriented towards the search for a balance between responsible rationality and instrumental rationality: the first of these, by identifying knowledge in a rational process, believes that it is possible to tone down the absolutising claims of the aspiration to objectivity in order to bring them back within the limits of both critical and self-critical caution that takes into account the ethical aspects implicit in every evaluation; the second, on the other hand, considering the knowledge that derives from research as a transformative process of reality, promotes the use of the most effective means to achieve the re-educational and treatment purposes that underlie every intervention project in the psycho-criminological field, ensuring - moreover - to act with integrity and ethical-deontological transparency.

The purpose precisely underlying not only this research, but above all the proposed application (and introduction) of these instruments in criminal justice is undoubtedly founded on the fundamental interest in translating scientific knowledge into activities that are useful and can act as support for professionals who have to express themselves with regard to the risk that a person may re-offend or re-offend. Risk assessment - in fact - does not aim to reduce criminal persistence or violent recidivism, but to assess the risk - in supporting legal practitioners - that a new criminal, anti-social or violent behaviour will be enacted. Therefore, risk assessment follows and makes use of certain criteria and fundamental aspects that must be taken into account when assessing, firstly, whether a certain person is at risk, secondly, in identifying what kind of risk we are talking about in order to be useful in establishing the most appropriate treatment or instrument to be used for the specific situation:

1. Predictive accuracy;
2. Predictive significance in defining the nature, severity of the risk;
3. Clinical applicability of risk assessment in the process of intervention planning and treatment orientation;
4. Level of internal reliability of the instrument.

3.2 *Brief hints on risk factors in risk assessments*

Within the risk assessment tools, there is a combination of various factors that intervene in the functioning of the tools and that constitute the so-called 'data processing'. The factors that are nothing more than the selective type of data that are 'entered' and then 'processed' by the instrument, are the derivative of a selective choice that is made at the basis.

While maintaining a purely legal perspective in the analysis of risk factors, it appears necessary, first, to provide a description of two elements that are closely related to risk assessment tools and that will prove useful in the following discussion of possible applications. In attempting to provide a description of what is meant by risk factors, it can be said that *'a risk factor can be any activity, lifestyle behavior, food, genetic predisposition, or environmental exposure that increases a person's chance of developing a disease'*. On closer inspection, the certainly most crucial point in the choice of application of these instruments concerns precisely the selection of which elements are to be considered 'risk factors' and which are not.

The basic premise, already mentioned in part, is that risk assessment is a term used to describe the overall process or method in which one:

- Identify hazards and risk factors that can potentially cause harm (hazard identification).
- Analysing and evaluating the risk associated with that hazard (risk analysis, and risk evaluation).
- Determine appropriate ways to eliminate the hazard or control the risk when it cannot be eliminated (risk control).

Risk assessment is a thorough examination of the workplace to identify elements, situations, processes, etc. that may cause harm, particularly to people. After identification, the probability and severity of the risk is analysed and assessed. Once this determination has been made, it is possible to decide what measures to take to effectively eliminate or control the harm.

- Risk assessment - the overall process of hazard identification, risk analysis and risk assessment.
- Hazard identification - the process of identifying, listing and characterising hazards.
- Risk analysis - a process of understanding the nature of hazards and determining the level of risk²⁸⁹.

²⁸⁹ CSA Standard Z1002 'Health and Safety at Work - Hazard identification and elimination and risk assessment and control' uses the following terms.

In conclusion, to the following introductory descriptions, the purpose of the risk assessment process is to assess hazards, then eliminate them or minimise the level of risk by adding control measures if necessary. By doing so, a safer and healthier workplace is created. Well, in the case in point, it should be noted that in the historical evolution that has led to the identification of various risk assessment tools, for the subject of discussion here, there are various types of factors, among which, following a macro distinction, it is possible to subdivide them into static factors and dynamic factors, which will be discussed in greater detail in the following pages.

The aim is to try to answer the following questions: what can happen and under what circumstances? What are the possible consequences? What is the probability of the possible consequences occurring? Is the risk effectively controlled or are further actions required?

After a historical description of the emergence and evolution of risk assessment, this work will attempt to analyse and select which categories of risk and which factors are to be taken into account in the specific case.

What is certainly relevant in the study on risk factors to be taken into account for risk assessment on reoffending is the fact that the practical application of risk assessment relates to the fact that the risk of reoffending or violent behaviour is not a static measure, but can vary due to different and multiple conditions (e.g. a successful treatment intervention or a change of context) or a different combination of individual elements, in the light of what is then called 'conditional risk'²⁹⁰.

Therefore, in conclusion to these premises, it is considered that an appropriate use of the risk assessment tool should be able to imply awareness of the interaction of these elements.

In fact, since risk always entails uncertainty, the possibility of identifying risk factors is certainly the first step to take if one wants to imagine a hypothetical introduction.

In doing so, it must undoubtedly also be noted that criminal behaviour does not result from single factors or inputs alone²⁹¹, but rather from the interaction of several factors that must necessarily be considered when assessing an individual's risk.

Therefore, the first step in the application analysis of these tools will be to identify the different factors that influence criminal behaviour.

²⁹⁰ On this point, see S. QUATTROCCOLO, *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale "predittiva"*, 1748.

²⁹¹ In the non-deterministic field of social sciences, it is preferable to avoid reference to 'causes', as it is more appropriate to use the term 'causal factor'; see D. FARRINGTON, *Early Development prevention of juvenile delinquency*, in *RSA Journal*, November 1994.

4 *The combination of static and dynamic factors*

In particular, risk assessment is a scientific discipline which, however, can only manifest itself as a 'concrete discipline'²⁹² that in its evolutionary path has moved from a purely intuitive and discretionary process of ideographic and impressionistic approximation, to a process of measurement and mere quantification of risk, to a process of integrated assessment in which the information and data collected follow a precise, rigorous and shared procedure, which gives incremental value to the evaluative analysis, since it combines methodological accuracy with clinical individualisation, coming to manifest or become a truly structured and integrated judgement. If one tries to select and group together, following the historical process that has led to the grouping and identification of various types of factors, one can see how a variety of factors could be identified.

It is therefore considered essential to 'decompose' the risk assessment into its different and sequential components, which must themselves follow the logic of the assessment:

1. Identifying risk factors
2. Measuring risk factors
3. Combine risk factors
4. Produce a final risk estimate/assessment to guide treatment.

Specularly, there are four different generations of risk assessment that have marked its scientific and applicative evolution, resulting in different steps in their transformation. There have been roughly four different generations of risk assessment tools over the course of the past century²⁹³. Indeed the four generations of risk assessment can be divided as follows:

- Generation I with a clinical and unstructured approach;
- Generation II with an actuarial or statistical approach;
- Generation III with a structured professional approach;
- Generation IV with a structured professional risk assessment and management approach.

Corresponding to these different categories is a parallel evolution of the risk assessment instruments which, in the first generation, for example, contemplated the element of social dangerousness as the focus of the analysis of the risk assessment instrument, to arrive then with

²⁹² G. ZARA, *Tra il probabile e il certo*, 59.

²⁹³ S. TURNER (et oths), *Development of the California Static Risk assessment (CSRA): Recidivism Risk Prediction in the California Department of Corrections and Rehabilitation*, Center for evidence-based corrections, University of California-Irvine, 2013.

the fourth-generation instruments at an integrated analysis in which the evaluative accuracy is associated with the evaluative specificity of the individual case for treatment and prevention purposes.

Starting with the third generation of risk assessment, these instruments have experienced a rapid spread in the United States, and this has led to the development and marketing of various systems, especially by private companies.

A further distinction could be made between instruments with an actuarial approach and those with a structured professional approach: the former consists of generalised mechanical assessments that assign a certain score based more on static risk factors.

The latter, on the other hand, formulate judgements that tend to be individualised, supervised by a technical advisor, taking into consideration a certain number of factors empirically and theoretically associated with the result that one is interested in having; there is a score for each factor whose relevance is then given by the technical advisor²⁹⁴.

Both approaches have their merits and drawbacks: the former generally offer objectivity, transparency and greater speed, but lead to a generalised and non-specific result; the latter, thanks to the discretion of the technical consultant on the final result, allow for the concreteness and specificity of the case submitted to them to be taken into account, but are more complex to manage, and human influence leads to a greater risk of reproducing bias in their outcomes. Both approaches have similar accuracy²⁹⁵.

We will now review the three main algorithmic risk assessment systems used in US jurisdictions; pointing out, however, that there are several other examples that could be considered.

Historically, the first form of specialist judgement of dangerousness was the unstructured clinical opinion, which still remains the only form among many national experts in the field. The opinion of the clinician and/or the treatment team becomes, in this approach, discretionary with regard to what kind of information or risk factors to take into account or omit for the formulation of a prediction of violent behaviour. Some pioneering studies in the 1970s came to the conclusion that risk prediction based on unstructured clinical judgement was only slightly superior to chance and was characterised by wide discretion among assessors. It was estimated that in only one out of three cases was unstructured clinical judgment valid in terms of predictive accuracy of future violent behaviour. The need to compare clinical judgment with

²⁹⁴ S. DESMARAIS - J.P SINGH., *Risk Assessment Instruments Validated and Implemented in Correctional Settings in the United States 2*, in *The Council of State. Governments (CSG) Justice Center*, 5 – 7, 2013.

²⁹⁵ *Ibidem*.

risk factors of statistically significant significance attracted the research to prospective studies aimed at cross-referencing a number of predictor variables with longitudinal observation of large cohorts of special patient populations. These variables were assigned a score that quantified the significance of frequency of occurrence and their correlation to the expressiveness of violent recidivism. The sum of the risk factors that make up the various psychometric scales developed was given the name 'actuarial', meaning the fixed, immutable, non-modifiable nature of the predictor variables, analogous to the methods by which insurance companies develop risk estimates with respect to a given event.

They take out life insurance, for example, on the basis of fixed risk factors such as age, gender, cigarette smoking, place of residence, etc., the statistical association of which produces a predictive estimate for each policyholder. Actuarial scales, unlike most psychological tests, are neither descriptive nor diagnostic; they aim to perform a predictive or prognostic function and are developed to predict a future event. The results of the actuarial scales are to be interpreted inductively as is the case in the following example: in the sample used to construct the actuarial scale, 56% of the persons in scoring X are known to have reoffended violent behaviour; Mr Y on the actuarial test is in scoring X: Mr Y therefore has a risk of reoffending violent behaviour similar to the percentage risk in the sample-population. Of the actuarial scales developed over the past two decades, the Violence Risk Appraisal Guide 16 (VRAG) is the best known and most widely used. Developed by prospectively studying a cohort of 600 subjects discharged from a Canadian forensic psychiatric hospital, it includes risk factors identified from among the multitude of variables represented in the sample that best correlated with the outcome variable violent recidivism. At the end of the 7-year follow-up, 12 variables were identified that were statistically significantly associated with violent behaviour, albeit with different 'weights': school problems, PCL-R score, personality disorder, alcohol abuse, separation from parents before the age of 16, parole failures, history of non-violent offences, never married, schizophrenia, previous, victimisation, age, female victim(s)²⁹⁶.

²⁹⁶ From the original sample from which it was developed, the VRAG predicted with an accuracy of 0.76 at the AUC violent behaviour at 3.5-year follow-up, 0.74 at 6-year follow-up, and 0.74 at 10-year follow-up 11, in a population of male, recidivist offenders described as 'serious'. The VRAG has shown good predictive abilities in different correctional and clinical settings, in sex offenders 17 18, in samples of forensic psychiatric patients 19, in the predictivity of recidivism in prison populations 20 21. In studies that have examined the validity of the scale between genders, differences in predictive ability emerge. Hastings et al. 22 followed a sample of about 500 inmates in a prison environment for a year and found that the VRAG shows significant predictive power both in oppositional and rule-breaking behaviour within the institution and in the probability of post-release recidivism in male inmates but not among females. Evidence of the latter is also confirmed by Coid et al. 23, again on prison populations, who report a significant predictive capacity of the actuarial scales used in the male sample, but lower than others, HCR20 and PCL-R, among women. The prospective study of a long follow-up, 11 years, by Kroner et al. 24 shows that about 38 per cent of the sample of 136 subjects discharged from a German judicial psychiatric

The nature of actuarial ERA scales, due to their characteristics of fixity and non-discretionary nature, to the extent that they can be administered by non-healthcare professionals, has led to a careful consideration of their limitations and possible margins of applicability. According to Hart 28 the limitations of an actuarial scale are:

1. focus on a small number of risk factors, it does not include potential case-specific factors;
2. the risk factors included are static, immutable and make the clinician's prediction a de facto 'passive' operation;
3. excludes those factors that have not found an empirical basis for correlation with the events to be predicted;
4. are tools developed to best predict certain adverse events in a given time period, in a given target population.

In a subsequent paper, Hart et al. 29 in assessing the margins of error at the group and individual level of the VRAG and Static-99 actuarial scales, analysed by calculating the 95% CIs, conclude that "the two scales analysed have poor risk prediction accuracy. The margins of error are substantial at the group level. At the level of the individual, the margins of error are so high as to render the test virtually meaningless'. The Authors' hypothesis regarding group predictivity rests on the margin of error that the sample selected for the construction of the scale is actually representative of the entire population in relation to those specific pathology and deviance characteristics (by way of example, VRAG includes the diagnosis of

hospital relapsed violent behaviour (mean time to relapse: 58 months) and that the VRAG applied to the same sample has a high predictive accuracy (AUC: 0.73). Evidence on the predictive ability of VRAG in non-forensic psychiatric contexts is more uncertain. The instrument developers retrospectively applied an incomplete version of the 10-item scale to the sample participating in the MacArthur study (n = 741) for a follow-up period of 20 weeks 25. The results indicate a significant ability of the scale to predict the number of violent adverse events and their severity according to the AUC method (0.72) and Pearson's correlations. At the same sample and on the same outcome indicators, Edens et al. 26 disaggregated the values of the 10-item VRAG, as applied, with the VRAG lacking the item corresponding to the PCL:SV (Psychopathy Checklist, Screening Version), and the single values of the PCL:SV itself. At the same 20-week follow-up, the 10-item VRAG showed a predictivity of violent behaviour of 0.73, thus similar to the work of Harris et al. 25, but the PCL:SV alone achieved an AUC of 0.78, which dropped to 0.64 for the VRAG without the psychopathy item. The authors derive from this that the predictivity of VRAG depends mainly on the psychopathy dimension. They also deduce that personality traits, in particular those generally afferent to an "antagonistic" attitude, as described by PCL:SV factor 2, assume significance as a robust correlate of potential violent behaviour. The authors downgrade Harris 25 conclusions that evidence of the ability of actuarial systems to predict violent behaviour in non-forensic psychiatric populations can be inferred from the data collected on the MacArthur sample. Grann 27 compared the ten historical items of the HCR-20, H 10, and VRAG in predicting violent recidivism 2 years after discharge from forensic facilities of 293 patients with a principal diagnosis of personality disorder and 111 patients with schizophrenia. Both scales proved to be good predictors of violent recidivism particularly in the sample of subjects with personality disorder, suggesting that the static/anamnestic items may better intercept the most significant recidivism variables for this category of patients, while the clinical and management/risk management items would be good predictors of recidivism among subjects with *schizophrenia*.

schizophrenia among the protective factors for violent reoffending). The confidence intervals widen since in biomedical research (as well as in the insurance industry) the sample sizes from which factorial grids are constructed are of the order of tens, sometimes hundreds of thousands as opposed to the little more than 600 individuals involved in the prospective validation study. For the authors, the effects on individual evaluation are even more pronounced and are illustrated by a game analogy²⁹⁷. The authors conclude by recommending extreme caution in the use of actuarial instruments, which should be confined to second-order decisions such as "administrative evaluations concerning the frequency and intensity with which specific risk management strategies are to be implemented with respect to the individual case". The limitations of actuarial scales lie mainly in the limited usefulness they offer in the specific clinical risk situation of violent behaviour represented by the individual case. It consists of deductive clinical-anamnestic indices, current psychopathological elements, environmental variables and the methodology and intervention strategies of the treatment team. An instrument that merely predicts a given risk without providing operational tools for the caregiver to monitor and prepare effective prophylactic interventions is of partial practical utility and raises legitimate ethical dilemmas for the clinician. The development of the third generation of ERA aims to synthesise the characteristics of the two previous methodologies by combining static clinical-anamnestic elements with empirical findings from clinical practice.

The design in this case is directed towards the development of a structured, evidence-based judgement on the prevention of violent behaviour, which is aimed at the management of possible recidivism and which leaves room for reflection and ultimately decision-making by the clinician, i.e. supports his/her choices²⁹⁸.

²⁹⁷ For example, we start from the assumption that a player has 3 out of four signs available to beat the dealer in the card game. If the two played 10,000 games we should expect a 75%-win rate from the player with a low margin of error, given the high number of bets (IC 95% of 74-76%). But if the number of bets decreases, the margin of error increases: IC 72-78% for 1,000 plays, 66-82% for 100 plays and 12-99% for the single play. If the gambler is the patient to be statistically evaluated with respect to whether or not he or she belongs to a behavioural risk category, the magnitude of the error margins may render the test performed of little predictive value.

²⁹⁸ The internationally established structured clinical assessment scale, widely used in forensic psychiatry in the Anglo-Saxon and Northern European area is HCR-20 (Historical, Clinical Risk, Webster et al. 30). The 3-letter acronym implies the presence of 10 items dedicated to the historical profile (H) of the patient, investigated in behavioural, psychopathological, personological and criminological aspects. The 5 clinical items (C) assess the patient's current condition, symptoms, conduct, insight and treatment compliance. The management of possible future risk situations (R), risk management, is investigated in the remaining 5 items, which include the viability of current and future therapeutic projects, the presence of potential stresses in the patient's environment, the availability of support figures in the patient's life. Since the appearance of the scale in the second half of the 1990s, a great deal of work has been produced in the literature to test its predictive validity and reliability, particularly in prison and forensic psychiatric settings (for an exhaustive literature review. The AUC values in the predictive validity studies taken together and for each scale administration setting are shown below (Tables I-III). In the first study on civil psychiatric patients, see K. S. DOUGLAS (et oths), *Relevance to violence risk assessment and*

4.1 *A deeper look: the different types of risk in relation to factors*

It should be noted here how different approaches or even risks have been 'isolated' as elements to be applied in the evaluation and processing of risk assessment tools. Therefore, the categories of risk that have been used the most and that have evolved in parallel over time will be analysed; in particular, these are: clinical risk, actuarial risk, structured occupational risk and then we will try to conclude by analysing the reasons why the latest generations of risk assessment are the ones whose application is proposed (and the possible and related advantages) and the most crucial moment of these tools, which is the identification of the risk in order to prevent its discounting.

It is relevant that the tools for implementing a practical, individual risk assessment are different, due to the different approach used in translating psycho-criminological theories into a personal assessment. Indeed, then the traditional main distinction is between actuarial instruments and clinical, professional instruments of judgement. In conclusion to this premise, it can be said that actuarial instruments use historical, static risk factors, whereas, on the contrary, professional assessment instruments take dynamic risk factors into account.

4.1.1 *The clinical risk*

On closer inspection, the first type of risk underlying the purely clinical approach (before declining into various forms and rarely used today) was nevertheless the most static approach. This type of approach led, in fact, to subjective, discretionary, impressionistic and static decisions since they remained anchored to the mere assessment of the expert. This element, no doubt, had obvious practical consequences, first of all the fact that it could not then be looked at in a broader and therefore generalised way and, secondly, that it remained conditioned by the expert's experience. Looking, therefore, at the intended use and objectives, it was felt that

manageme in Forensic conditional release contexts, in *Behavioral Sciences & the Law*, 2014, 32; followed a cohort of 193 subjects in psychiatric care with previous arrests for violent offences, predominantly male, 30-40 years of age and of Caucasian ethnicity, for a period of more than 2 years, comparing HCR-20 values with the short version of psychopathy, PCL:SV. The results at AUC showed a predictive ability of the structured scale between 0.76-0.80 for violent behaviour events vs. a range of 0.68-0.79 for PCL:SV. Cross-referencing the individual predictive abilities by means of logistic regression analysis, the authors observe that the 19 items of the HCR-20 increase the individual predictive abilities of the psychopathy scale, but not vice versa. Also in Europe, in the Anglo-Saxon area and northern countries, validation and evaluation studies of the instrument appeared with the beginning of the noughties, either alone or in combination with actuarial or structured clinical analogues. Doyle and Dolan 33 followed the performance of 112 subjects discharged from forensic and non-forensic residential facilities for 24 weeks. The results confirm that the HCR-20, but also the PCL:SV and VRAG, are significantly predictive of recidivism of violent behaviour of patients discharged to the territory. Some considerations of interest emerge from the work: VRAG is accurate in particular for patients discharged from forensic facilities but not for patients followed within returned pathways.

they would not be able to predict violence with reasonable scientific certainty for two primary reasons: firstly, the use of informal assessment and decision-making criteria that were not defined *a priori*; secondly, the reliance on certain specific characteristics of the patients, but which were generic because they were not associated with elements that related to a certain degree of criminal and violent persistence.

From the earliest use, the purely clinical and unstructured approach already possessed problems of systematicity and repeatability. In addition, another critical aspect that came to the fore early on was that clinical practitioners tended to ignore the base rate of violence (base rate) and many of their evaluations were therefore not based on specific reference populations but only on their own case histories of the patients examined. In fact, the clinical assessment did not provide for an identification of here risk factors or criminogenic processes that were specifically pointed out as being significant and on which to focus the analysis; what was instead considered relevant for clinical judgement were rare and atypical events that in themselves are highly infrequent and therefore also uninformative with regard to the functioning of the person in his or her living environment. In this context, a further issue of relevance concerned the fact that the unforeseen event, should it occur, could never be anticipated and could not be fully and completely assessed. Well then, since in any case (never, or almost never) one believes one can arrive at complete and absolute knowledge given one's limited cognitive capacities (limited rationality), one must instead succeed in making the best use of all the information available, thus attempting to be able to grasp the significance of events that can be anticipated, given certain conditions and factors, and knowing how to use only those relevant to the evaluative task (informational selectivity).

4.1.2 The actuarial risk

The actuarial judgement approach is based on prospective and retrospective longitudinal studies that aim to cross-reference a set of predictors with observations over time of large populations of persistent criminal individuals or psychiatric-forensic patients, or of violent individuals involved in heterogeneous or specialised criminal careers.

In particular, this type of approach is able to provide a probabilistic estimate of the risk of violence using an algorithmic procedure that assigns a score that quantifies the significance of the different risk factors observed longitudinally, measured in terms of frequency of occurrence and strength of correlation with criminal and violent recidivism, both general and specific. Furthermore, the actuarial summation of the risk factors composing the different psychometric scales refers to the historical, static, unchangeable nature of the predictor variables. In

particular, these actuarial assessments have always proved more accurate than the purely clinical method; they are also generalisable and offer a transparent description of the rules and method used to identify and measure the risk of criminal relapse. Here, the concept of accuracy is understood in terms of the specificity and sensitivity of the assessments as described above.

To date, various actuarial tools exist and are used, especially in the field of criminology and custody²⁹⁹. The level of predictive accuracy of such instruments in circulation today is mostly given by the use of static and historically relevant risk factors in the commission of persistent and violent criminal behaviour. In particular, these are assessments that are not diagnostic or descriptive of a person's mental, psychological or relational functioning, but instead allow an accurate estimate based on what is most likely to occur given the subject's past history.

It is now believed that the constant emergence and proliferation of instruments based purely on risk calculation is changing - or redesigning - the Anglo-Saxon 'penal field' in a wholly structural way, ushering in an era of actuarial justice. Indeed, the assessment methods used in the criminal field are gradually abandoning the traditional clinical approach, based on a psychological analysis of the subject's dangerousness, and moving to statistical methodologies that instead assess the riskiness of the actuarial category to which the defendant or offender belongs. Indeed, this 'shift' not only de-individualises the assessment process, but at the same time shifts the emphasis from the goal of re-educating offenders to the management or administration of individuals who are classified into various risk groups³⁰⁰.

Actuarial predictive accuracy continues, in fact, to be the highest, since, by assessing historical, static and unchangeable conditions, it is, without doubt, the most accurate. However,

²⁹⁹ There are many well-known and widely used actuarial tools. For example, the VRAG is a violence predictive instrument used for adult criminal individuals in custody, in psychiatric-forensic patients and non-forensic patients, also in the complex conditions involving sexual and domestic violence; in this case the predictive accuracy is very high thanks also to the integration of the evaluation of psychopathy, measured with the PCL-R. on the other hand, an application limitation of the VRAG has already been noted, which lies in the time required for its administration; in fact, it requires the collection of information related to the person's life history, also dating back to childhood, a psychiatric assessment and data on the criminal career. In particular, the PCL-R is the actuarial instrument constructed and developed to measure psychopathy in the prison, criminal, psychiatric- forensic, and civil population most widely used by the scientific community. In addition, a screening version and a version for juvenile individuals (the Static-99R and Static-2002) have also been developed and are widely used to measure the static risk of sexual violence. The LSI-R, also in its risk management version (LS/CMI) and in the version specifically constructed for juvenile individuals (YLS/CMI) allows instead to assess the risk of criminal relapse in criminal individuals involved in alternative measures to detention or in the care of social services.

³⁰⁰ In other words, the rehabilitative approach used, at least on paper, in the penal sphere since the 19th century lost legitimacy at the end of the 20th century. The collapse of the 'grand narrative of penal modernism' paved the way for retributive and neo-liberal policies involving greater penal severity and the adoption of risk-based approaches and techniques (8). Much of these concepts resonate with Michel Foucault's work in that they suggest that a transition from a disciplinary to a biopolitical approach is underway, i.e. the legal system is abandoning the goal of normalising individuals in favour of identifying and managing populations.

it seems appropriate to reiterate that predictive accuracy is incomplete information if it is not also followed by an assessment of those risk dimensions that are modifiable by intervention.

4.1.3 *The professional structured risk*

The structured professional judgement approach, on the other hand, moves in the direction of integrating evaluative accuracy with clinical thoroughness (sensitivity, integrity, ethicality), by means of a method that is precise and statistically valid. This approach would take into account criminogenic needs³⁰¹, which can be modified through a treatment intervention. This approach is also called an aide-memoire³⁰² as it offers a set of guidelines to conduct the assessment, identify specific risk factors, and then organise the intervention³⁰³.

Now, this last generation of risk assessment is gradually being joined by a fourth one that associates risk assessment with risk formulation and management, where the level of criminogenic needs is flanked by the assessment of protective and promotive processes, and of compliance, offering the expert operational tools to monitor and prepare therapeutic and treatment interventions. In this space of flanking, interdisciplinary and inter-professional integration is the one that would seem to best realise the constitutional principles of social protection and promotion of the health of the individual and the community, of active and humanised empowerment of the criminal individual, whether or not chargeable, of re-educative opportunities of punishment and social reintegration.

Indeed, through the systematic collection of coherent and necessary information, the use of an appropriate methodology, and the use of specific risk assessment tools, one avoids significant variables escaping scientific attention or unspecific variables being included in it, leading to inaccurate assessments that are more frequent the more the expert operates in an emergency or the higher the level of professional stress and emotional and cognitive overload. Professional and structured risk assessment is a valid procedure as it is based on scientific research.

In addition, a further aspect that characterises this procedure is that it refers to the concept of *conditional judgement* with respect to the assessment of the risk of violence, since the manifestation of violence only occurs given particular conditions of the individual's life and not independently of them.

³⁰¹ *Segue*, § 5.2.1.

³⁰² G. ZARA, *Tra il probabile e il certo*, 66.

³⁰³ For example, the HACR-20, mentioned in the previous paragraph, in its third edition is an example of this type of instrument.

However, it should be noted that, however much any assessment aims to achieve an optimum degree of accuracy, this will never be absolute, considering the variability inherent in people's lives, variability that cannot always be anticipated in the way it manifests itself. Indeed, the possibility of identifying a complete set of 'conditions' would render any risk null and void: while theoretically this constitutes an ideal, in reality, it is impossible. On the other hand, it would be like expecting the environment to adapt to the conditions of each individual person, in order to reduce the risks of criminal persistence; this, however, would mean guaranteeing a kind of absolute prediction of feasibility in every assessment. This would imply that risk assessment must not be detached from the psycho-social reality of the person and the community in which he or she lives. It follows that changes are conditioned by the context and by what it is able to offer and support at a given and precise moment in relation to a specific individual, in relation to a specific request for intervention planning³⁰⁴.

Conditions and specifications were also discussed, highlighting and reiterating the fundamental importance of contextualising the risk assessment to the psychosocial reality of the individual under observation. In this case, it should be specified that by conditions are meant those characteristics of the person and his or her life that may support criminal re-offending (such as, for example, certain factors in particular, such as alcohol and/or drug use, pro-criminal attitudes).

Next, by specifications are meant those aspects of the criminal and violent incident that go beyond the mere occurrence/non-occurrence of the act itself and instead refer to specific characteristics of the behaviour, such as the type of criminal act that is committed, the target at which the violence is directed, the victims involved, the location of a specific criminal event where the act occurs. Undoubtedly, making a prediction regarding the specification is particularly difficult and would be less likely than a general prediction of criminal behaviour. For instance, it is considered more complex to identify the specific target or victim of a violent

³⁰⁴ Some authors also spoke of conditions and specifications, emphasising the importance of contextualising the risk assessment to the psychosocial reality of the individual under observation. One speaks of conditions meaning all those characteristics of the person and his or her life that may support criminal re-offending (such as: unemployment, alcohol and/or drug use, pro-criminal attitudes). On the other hand, we speak of specifications as meaning those aspects of the criminal and violent incident that go beyond the mere occurrence/non-occurrence of the act itself, but instead refer to specific characteristics of the behaviour, such as the type of criminal act that is then committed, the target at which the violence is directed, the type of victims that are involved, the place where a criminal event may occur, and the period of occurrence of the act after the individual has been released. It is believed that a prediction concerning the specification is particularly complex, but at the same time less likely than a general prediction of criminal behaviour. It is, for instance, very difficult to identify the specific target or victim of a violent act, compared to predicting that an antisocial individual will reoffend in the future. In fact, the prediction of criminal specification is certainly the most problematic because while it is possible to identify the probability of criminal relapse in an individual, it is much more difficult to identify the type of crime he or she will commit and at what time, the manner of perpetration and the victim involved.

act than to predict that an antisocial individual will commit a criminal act in the future. Indeed, the prediction of criminal specification is certainly the one that poses the most problems because, while it is possible to identify the probability of an individual's criminal relapse, it is much more difficult to identify the type of crime he or she will commit and at what time, the manner of perpetration and the victim involved.

5 *A crucial step: risk identification*

A fundamental aspect in the use of such systems concerns the identification of risk. In fact, the counting of the risk factors and the weight assigned to them does not exhaust the risk assessment, just as the actuarial nature of the risk assessment must not, for its part, replace the role that the expert can play in clinical observation, in interviewing the person, in formulating an integrated judgement, in formulating and planning an inter-professional and inter-institutional strategy, which is adherent to the purposes of his mandate and, above all, consistent with the person's needs, his resources and the possibilities of treatability.

Preliminarily, it should be noted that the term risk refers to the probability of an event to occur; the term risk factor refers, on the other hand, to that condition or correlate that precedes an outcome and is a condition that implies an increased probability that a criminal event will occur, that it may reoccur under the same terms or in a different way from how it occurred in the past.

The key to the distinction between *correlatum* and risk factor lies in the so-called temporal precedence, since a risk factor is that specific condition which temporally anticipates an event, whereas a correlatum is a factor which is associated with the event, does not anticipate it and thus represents a symptom or indicator of the outcome. Therefore, all risk factors can be understood in terms of correlates, but not all correlates are risk factors.

What is crucial to a full understanding of how these instruments work is that not all risk factors are the same and not all impact unequivocally in the same way on the outcome then issued. On closer inspection, risk has its origin in biology, psychology, psychopathology, family and culture. At the same time, it also has its own temporality, in the sense that some risk factors have a significant influence on the individual's behaviour in adolescence, such as peer group pressure; others, however, become significant in adulthood, such as substance addiction or easy access to weapons or easy contact with potential victims; still others are pervasive throughout life, such as antisocial personality disorders or distorted pro-criminal thinking.

Basically, risk also possesses a dynamism of its own, in the sense that there are static, non-modifiable factors, stable dynamic factors (such as, for example, traits of impulsiveness and hostility) that are instead modifiable by intervention; acute risk factors that instead change and mutate rapidly and are associated with a condition that facilitates violent reactions (an emblematic example, in this case, is drug use).

If, in fact, the assessment does not result in a treatment that adequately meets the dynamic criminogenic needs, such an assessment will be as useless as ever and as a pure statistical exercise. For this reason, the reconstruction of the state risk of the individual is central in the risk assessment and in the construction of possible scenarios of future criminal and violent behaviour (the so-called risk formulation), in which that specific offender is most likely to 're-offend' or react in a violent or criminal manner.

The integration of risk assessment, risk management and risk reduction, as proposed in the third and fourth generation of risk assessment, requires a privileged focus on what is also referred to as the dynamism of risk³⁰⁵. Indeed, the concept of dynamism is influenced by the nature and temporality of risk. There are factors that are pervasive (e.g. personality disorders); others are context-dependent (e.g. accessibility of weapons or ease of contact with the victim); while others may be accentuated by life experiences (e.g. distorted or pro-criminal thinking).

Risk processes change through intervention over time depending on whether they are dynamic stable or acute. The division of risk factors according to their dependence on the time dimension is certainly a simplification.

The *risk status*, on the other hand, refers to the identification of differences between different groups of individuals at risk; it also highlights differences between individuals (between individual differences); it specifies the risk status in one individual with respect to another or between different groups of individuals. Risk state, on the other hand, implies the intra-individual level of risk of an individual in a particular condition or moment of his or her life and, furthermore, the fluctuation of the individual disposition to commit violence depending on the biological, neuropsychological, psychological, relational and cultural components that condition an individual's choices and behaviour. The understanding of what constitutes risk of violence and criminal persistence, of what is the temporality, statistic and dynamism of risk therefore becomes extremely relevant and it is the epistemology of risk that will form the basis of the later reflection.

³⁰⁵ G. ZARA – D. FARRINGTON, *Criminal recidivism: explanation prediction and prevention*, 16.

In the analysis to be carried out in the following pages on the differences between the two macro-categories of factors (static/historical and dynamic), it is already anticipated from the outset that there is no evidence that static or dynamic risk factors are more or less reliable indicators of risk. Indeed, in themselves, both are insufficient and must be considered together in order to make a correct assessment. However, it is already anticipated here that it is not possible to imagine 'constructing' and 'training a tool' that is capable of assessing all risk factors; that is, that it is capable of 'covering' every potential protective factor³⁰⁶.

5.1 *Static and dynamic risk factors*

These are, in particular, those unchangeable factors (e.g. previous criminal career, age of initiation or number of convictions) that are robust predictors of future behaviour, on which, however, no intervention is possible, as they cannot be changed. Static risk factors have been defined as fixed risk markers, in the sense that they contribute to criminogenic influences and, while constituting areas of treatability, they become informative dimensions for organising intervention: they identify, in fact, criminal individuals at high risk of violence and criminality on whom treatment and clinical attention is a priority. These include gender (being male or female), race and genotype. Indeed, modifiable risk factors are further differentiated into variable risk markers and causal risk factors or criminogenic needs. The former is always modifiable, but it has not so far been demonstrated that a modification of the latter is directly associated with a reduction in the risk of criminal and violent behaviour.

5.2 *The dynamic risk factors*

After illustrating the static risk factors, we see how the dynamic risk factors, on the contrary, can change according to the individual's situation. They include so-called criminogenic needs³⁰⁷, i.e. aspects of a person or his situation that, when altered, may imply a change in his criminal behaviour³⁰⁸.

5.2.1 *The criminogenic needs*

By criminogenic needs we mean those dynamic psychological risk factors that are directly linked to antisocial initiation or that can contribute, by mediating it, to criminal continuity.

³⁰⁶ 'Undoubtedly, the integration of different sources of information is crucial to ensure more reliable results,' see S. QUATTROCOLO, *Sui rapporti tra pena, prevenzione del reato e prova nell'era dei modelli computazionali psico-criminologici*, in *Mimesisjournal*, 271.

³⁰⁷ *Segue* § 4.2.1.

³⁰⁸ See S. QUATTROCOLO, *Sui rapporti tra pena, prevenzione del reato e prova nell'era dei modelli computazionali psico-criminologici*, in *Mimesisjournal*, 271.

They are, in particular, modifiable factors on which intervention can and must be planned to promote a reduction in the risk of criminal initiation and continuity. Studies are homogeneous in considering anti-social personality characteristics, an anti-social family and relationship network, pro-criminal attitudes, alcohol and drug abuse and dependence, and anti-social recreational activities as criminogenic needs. Indeed, adequate consideration of criminogenic needs can undoubtedly facilitate intervention planning and foster a higher level of intervention responsiveness.

Criminogenic needs are dynamic and dimensional psychological risk factors that refer to characteristics of the person involved in a persistent career and his or her life situation that can be modified by the intervention. Once they are modified, they promote a significant change in antisocial potential³⁰⁹, behavioural externalisation and criminal career. Criminogenic needs involve at least eight dimensions of the individual's life. In particular:

1. Anti-social history and criminal career;
2. Personality disorders, aversive emotionality, lack of self-control, impulsivity;
3. Distorted thinking, pro-criminal attitudes and cognitions;
4. Pro-criminal and antisocial network;
5. Inadequate and distressed family conditions and/or conflicting and problematic marital situation;
6. Substance dependency;
7. Recurrent problems in the school or work context;
8. Lack of prosocial recreational activities.

5.2.2 *Psychosocial needs*

Within the same category of dynamic risk factors are psychosocial needs, which are those risk factors present in the person's life reality and which contribute to altering his relational and social functioning. They are not directly associated with criminal behaviour and its continuity, but they influence his or her adaptation to the environment with respect to social demands and standards, making him or her more vulnerable to antisocial pressures.

Indeed, the strength of the association between criminogenic conditions and the likelihood of violence and persistent criminal manifestations is directly proportional to the number of risk

³⁰⁹ D. A. ANDREWS – J. BONTA, *Rehabilitating criminal justice policy and practice*, in *Psychology, Public policy and Law*, 16(1) 2010, 39-55.

factors involved; in fact, the more numerous the risk factors, the higher the likelihood of violent outcomes (this result corresponds to the principle of summativity of risk factors). However, the aforementioned principle does not only refer to a quantitative and linear issue, whereby in the presence of 2,5,9 risk factors the probability automatically doubles or quintuples. This underlying risk principle is that of the dose-exposure relationship: the precocity, duration and intensity of exposure to several risk factors interacting in a cumulative, equifinal, dynamic manner and increasing the likelihood of violence and criminal manifestations.

In order to be truly effective, assessments should be able to go beyond the identification of inter-individual variability of risk between individuals (between individual differences). This difference tends to remain constant over time: criminal individuals at low risk of violence will always continue to differ from criminal individuals at high risk.

What assessments should instead focus on is a greater clinical attention to intra-individual variability (within individual differences) in the potential for violence that changes over time. This last differentiation, on the other hand, has led scholars to distinguish two risk assessment models: one oriented towards predicting the risk of violence and criminal relapse; the other, instead, aimed at violence reduction. The first is anchored in the certainty of the stability of the static risk (status risk), which is accurate to the extent that, being unchangeable, it traces a state of affairs. The second, on the other hand, is the one that implies the processuality of risk assessment and requires the identification of dynamic causal risk factors or criminogenic needs (the state risk) that when modified then activate changes in the outcome.

It should be noted here how two American scholars have made a further analysis leading to the identification of another specificity, identifying the dynamic risk factors to then move from an accurate assessment of the level of risk to an appropriate risk reduction in the case of psychiatric-forensic individuals. Many of these new factors actually correspond to other factors that have also been found to be important in the persistent, non-psychiatric criminal population.

Indeed, the results of clinical research show that many of the treatment programmes with mentally disordered offenders, oriented mainly towards the mere reduction of symptoms, have been ineffective³¹⁰. Indeed, many disorders are maladaptive, but the nature of the intervention must not be reduced to a mere elimination of the symptom, but must start from an acquisition of knowledge on the long-term course and on those factors that contribute to possible variations

³¹⁰ On this point, a U. FORNARI, *Al di là di ogni ragionevole dubbio. Ovvero sulla cosiddetta prova scientifica nelle discipline psicotforensi*. Turin, 2012.

in order to understand whether, and in what terms, a psychopathological improvement can contribute to an improvement in the person's social functioning.

In fact, it remains to be considered that personality disorders are also conditions of internal experience of behaviour, serious and persistent, leading to a functional and relational impoverishment of the person. Indeed, intervening to reduce the impact that criminogenic needs have in triggering violence in persons suffering from personality disorder undoubtedly has a more sustainable preventive objective.

6 *The I.N.U.S. conditions of the criminal behaviour*

First of all, it is necessary to dwell on the concept of 'cause', meaning that condition which has inherent in it the concept of effect; after all, modern medical, psychiatric and psychological sciences operate according to a probabilistic model since it is impossible to isolate a single cause of a problem.

Indeed, as already mentioned, probability theory defines the relationship that exists between risk factors and effects in terms of an increase in the probability of the effects in the presence of the examined risk factor and is the one that can best be applied to the study of criminal behaviour and its consequent evaluation³¹¹.

A cause, therefore, is considered effective when it becomes the means by which an effect then occurs. Mackie's model is used to present a useful interpretation of the causation behind human behaviour such as criminal behaviour, in which a series of chain events follow one another and the identification of a precise condition from which other conditions emerge is a matter of choice, the certainty of which is conditioned by the context, the effects, and the influences of other concomitant conditions.

In a criminological context such as the one described above, it seems appropriate and pertinent to introduce the concept and use of the INUS condition: indeed, a risk factor or criminogenic need such as antisocial personality disorder is an INUS condition of criminal persistence if the personality factor is an insufficient but necessary component of a series of conditions that together are not necessary, but have become sufficient for the commission of the offence. Trying to enter the subject in more depth, one tries to describe what is meant by

³¹¹ Consider, for instance, the probability that a certain factor is associated with the presence of criminal behaviour. In this case it is therefore assumed that a risk factor may contribute (partially or totally) in causing the event; the attribution of the latter to the former must imply a temporal relationship, whereby the factor precedes the effect; the identification of possible covariations with other factors (so-called confounders) can be detected by means of multivariate analyses, which make it possible to measure the presence and strength of covariations on the effect in their presence or in their partial and sequential removal.

this risk factor: an example can be given if one thinks of a condition of extreme impulsiveness, of poor self-control, of high aggressiveness: in such a case one is faced with an insufficient but necessary part of a condition that in itself would not be necessary, but which is nevertheless sufficient in the pattern of antisocial personality disorder. Thus, antisocial personality disorder is an INUS condition of persistent criminal behaviour.

In summary, therefore, risk factors, as well as criminogenic needs, are all INUS conditions that have the following characteristics:

- Summativity or cumulatitivity. This means that the more factors and needs that are present in a given life condition, the higher the risk that they may influence the person's life, since their impact is not given by a simple accumulation of different factors, but by the relationship that some factors have with others, increasing or attenuating their influence (positive or negative);

- Specificity: where certain criminogenic needs and specific risk factors act in a differentiated manner, under certain conditions, on certain individuals and not on others.

- Temporality: in the case where certain risk factors, which have a significant criminogenic incidence in a particular period of development, may certainly have no effect or a reduced effect in another period of life;

- Sequentiality: where certain criminogenic factors and needs appear to operate according to sequential behavioural sequences, almost preparatory to delinquent and violent behaviour;

- Equifinality: where different initial risk conditions may lead to the same behavioural outcome;

- Multifinality: when the same risk factors act differently, in different situations and with different individuals. One and the same existential condition may indeed finalise in a variety of successive, diverse and distinctive conditions.

In conclusion, in the psycho-criminological context there are neither sufficient nor necessary conditions for determining criminal behaviour, only INUS conditions.

6.1 A case in point: the Risk Need Responsivity model

In the criminological context, the risk need responsivity model is the one that best responds to the principles of individualised, targeted, reintegrative intervention, which we believe can also find a space in the Italian penal context.

An attempt will therefore be made to describe it. In this model

- the risk ® is to identify who to treat and to adapt the level of intervention and treatment to the level of risk

- Criminogenic Needs (N) is about understanding what to treat since the targets of treatment must be criminogenic needs;

- Responsiveness ® is about understanding how to treat in order to achieve treatment adherence which implies temporisation and specificity.

Specifically, the principle of responsiveness or responsivity outlines how re-educational and social reintegration programmes should be organised to converge with the individual's cognitive personality, emotional and socio-cultural characteristics and protective resources. On the other hand, compliance implies motivation on the part of the individual to pursue a treatment programme, active participation of the individual, interest in change and, ultimately, maintenance of the choice.

Since it is a 'biphasic/dual process', responsiveness necessitates the acceptance of the intervention setting, in which the person's time and readiness for treatment must be taken into account and then become synchronised.

On closer inspection, it is precisely the complexity of prognostic judgement that is well summarised by the 'risk need responsiveness' model, which embraces all dimensions of a prediction of the offender's future behaviour in relation to his individual characteristics and readiness for treatment.

7 *The dual application front of risk assessment tools in criminal justice*

The large body of literature devoted to the subject of predictive risk assessment has given scientific validity to an aspect of forensic psychiatry to which few in the late 1980s recognised any evidence value. The focus on and study of static and later dynamic risk³¹² factors has produced a conceptual shift in clinical thinking away from the historical concept of dangerousness, a dichotomous legal concept that by its very nature lends itself poorly to evaluations of gradualness and in any case does not provide indications on the management of relapse prevention. This, in the literature, in the heuristic framing of the problem, has in fact been replaced with the concept of risk, relating to the probabilistic and statistical nature of the morbid phenomenon and therefore suitable for investigation, quantification, standardisation

³¹² Con questo concetto e quindi con fattore di rischio dinamico si intende «any factors that contribute to recidivism risk that can change over time». On this point, D. KEHL – P. GUO – S. KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing, Responsive Communities Initiative*, in Berkman Klein Center for Internet & Society, Harvard Law School, 2017, 9.

and critical review. Originating within the North American prison system, research on predictive risk tools has gradually expanded to forensic psychiatry and, more cautiously, to general psychiatry. At first glance, the usefulness of static-actuarial predictive assessments outside the prison and forensic psychiatry fields may not be apparent. Developed in the second half of the last century within the US justice system with specific outcome purposes on inmates who were being assessed for the appropriateness of parole or other forms of mitigated penal enforcement, they present, from a perspective of applicability to non-judicial psychiatric patients, some not inconsiderable ethical dilemmas. These scales may be perceived as instruments at risk of stigmatisation and social marginalisation of the patient due to their character of fixity with respect to the immutability of the anamnesis and to the dichotomous outcome that does not contemplate treatment and social inclusion hypotheses. For these reasons, their use is discouraged for judgments of primary importance, such as psychiatric-forensic evaluations in the court context, but they may find appropriate use in supporting the clinician's investigation in identifying elements of the patient's vulnerability to violent behaviour. Static factor analysis can provide an estimate of the long-term probability of aggressive behaviour, describing the so-called risk status of the patient; it provides the clinician with a structured collection of the patient's anamnestic data, helps him/her to avoid negative counter-transference reactions so frequent with this type of patient, and supports a systematic data collection that can serve as a complement to a complete and coherent clinical deductive process, in which inter-individual differences relating to risk, e.g. certain personality traits, receive adequate analysis. Alternatively, an approach that takes into account static variables may reduce the chances that a position of inappropriate 'complacency' arises towards certain patients.

The collection of actuarial information can help the clinician in the identification of those cases that need more monitoring and support, the possible preparation of treatment programmes at higher levels of containment and protection, and the activation of more timely and assertive intervention strategies in the face of clinical signs of decompensation or relapse. This type of information, as already mentioned, should not exhaust the task of analysing and preventing recidivism variables of violent behaviour. The mere inclusion of static variables to the deductive reasoning could induce a pernicious conviction of clinical staticity, propaedeutic to 'therapeutic nihilism' and to unjustified long-term restrictive measures. These variables must, as seen in the preceding paragraphs, be supplemented with the evaluation of dynamic situational factors, modifiable over time. Factors such as the current psychopathological picture, substance abuse, non-compliance with pharmacological treatment, environmental

stressors may in fact vary over time and correlate with the likelihood of violent behaviour, thus providing an estimate of the short-term probability, the patient's state risk. The structured clinical approach to risk assessment substantiates the difference from second-generation instruments in the non-binding indication of clinical opinion, supports it in a systematic collection of the necessary information, and provides indications for the formulation of a judgment that is analogical and descriptive of risk, low-medium-high. As Maden warns, the counting of the risk factors does not coincide with the risk assessment, but it is a premise for the elaboration of an integrated judgement and of a planning strategy by the multi-professional care team: the structured clinical assessment of the risk of violent recidivism has no significant clinical usefulness if it is detached from a reflection on the strategy of prevention of recidivism, which is articulated in the phases of the description of possible future scenarios of violent behaviour, the risk formulation, synthesis of the careful anamnestic, clinical, criminological and psychodynamic analysis of the past events. Although the scenarios of future violent behaviour can be potentially unlimited for an individual who has already committed violent acts in the past, clinical observation shows that for most patients these do not exceed two to three possible alternatives and that in most cases they are limited to one. The description of possible scenarios is mainly based on the assessment of the patient's violent history and the possibility of its recurrence. The possibility of the unexpected event, not preceded by premonitory signs, occurring, however, belongs to the experience of every clinician and is also highlighted by retrospective studies on psychiatric homicide patients. Reflection on future risk assessment scenarios, according to the structured clinical model, far from being a kind of 'magical' practice of predicting future behaviour, is more wisely aimed at containing phenomena of patient malfunctioning that have already become apparent and are likely to re-occur. These eventualities weigh relatively few but highly frequent variables, such as pharmacological non-compliance, disease relapses and substance intoxications, on which management analysis must necessarily focus. The ERA tools, which are the most widely used in the international community and the most reliable, should support and make this clinical process as structured as possible, acting as tools of assistance and 'decision support' to the psychiatrist. They allow for systematic and methodologically consistent data collection, preventing significant variables from escaping the inclusion of the assessment. The structure of data collection also aims to avoid assessment errors, the 'heuristic biases', which are all the more frequent in the case of cognitive overload. Information and evaluations that are not primarily related to the ERA are thus kept in the background and are less likely to influence the evaluation. An evidence-based ERA facilitates exchange and communication with other

members of the care team, the patient and his or her family and affective network, making clear the clinical framing paths leading to a given assessment and any points of disagreement.

Attempts to structure approaches to the assessment and management of the recurrence of violent behaviour in psychiatric patients have been accompanied since their emergence by ethical dilemmas and perplexities as to whether they should be used. If on the one hand the phenomenon of aggression and heterodirected violence is included in the phenomenal manifestations of the psychic disorder and as such is the object of reflection and, if possible, of prevention by the caregivers, on the other hand the current state of scientific knowledge makes instruments available to clinicians with still very high false positive and partly false negative rates. The VRAG, the most accurate instrument currently available, has a sensitivity of 73% and a specificity of 63%, below what is considered acceptable in medicine for a screening instrument (chest X-ray is not used as a screening for pulmonary K because it has 'only' a sensitivity of 83% and a specificity of 90%) . It follows that the clinician, in addition to considering it far removed from his or her training and professional identity to reflect on any form of social control through the elaboration of 'predictive' hypotheses of dangerous hetero-directed behaviour, is often reluctant to use instruments that may induce erroneous judgements in patients who are potentially interned for no reason (false positives) or free but dangerous (false negatives). The ERA scales have all been developed to minimise the possibility of false negatives, leaving room for an assessment that may contain false positives. The sources of error and uncertainty associated with structured risk assessment impose further efforts on scientific research to develop more accurate and efficient ERA tools. In the meantime, however, it seems to us appropriate to point out that for patients who have already perpetrated violent acts in the past, 'difficult' patients whose problems of impulse control, comorbidity with addictive behaviour, risky lifestyle, and for those who report thoughts of threatening aggression towards third parties, a structured ERA assessment is considered in the international literature to be the 'standard of care'.

It should be noted at the outset that this paragraph will serve as a premise for the analysis and proposals that will be developed in the remainder of the paper.

With regard to the objective pursued with the recourse to algorithmic risk assessment systems during the various stages of criminal proceedings, it is a matter of balancing a twofold need: on the one hand, the protection of security and public order and, on the other, the defendant's rights of liberty. In particular, risk assessment tools can be used to: 1) assessing the possible existence of conditions for maintaining pre-trial detention when provided for (pretrial

risk assessment instruments-PRAIs); 2) assessing the risk of recidivism or the admissibility of alternative measures to detention (risk assessment instruments-RAIs).

In both cases, therefore, the analysis, at the conclusion of this paragraph, will bifurcate and follow the two paths in which it is deemed appropriate to propose the application use of these tools.

7.1 *The application paradigm of risk factors*

With regard to the objective pursued with the use of algorithmic risk assessment systems during the various stages of criminal proceedings, it is a matter of balancing a twofold need: on the one hand, the protection of security and public order and, on the other, the defendant's rights of liberty. In particular, risk assessment tools can be used to: 1) assessing the possible existence of conditions to maintain pre-trial detention when provided for (pretrial risk assessment instruments-PRAIs); 2) assessing the risk of recidivism or the admissibility of alternative measures to detention (risk assessment instruments-RAIs) with a view to an ad hoc choice on a personalised treatment of offenders; 3) in the application of measures that provide for a risk assessment to be imposed in specific cases in security measures.

At this first stage, we will limit ourselves to a brief description of the possible uses related to their characteristics. At a later stage, in the continuation of the paper, an attempt will be made to analyse the possible application implications following the twofold line: from a preventive point of view in the assessment of the risk of dangerousness and, secondly, in the application of such instruments to identify the best sanctioning treatment.

To yearn for a 'just' sentence or an 'exact' prognostic assessment of criminal dangerousness seems to constitute not only the wish of those involved in criminal proceedings as defendants, but the ambitious goal of the human being who ceaselessly strives to overcome boundaries that are moved ever further.

Indeed, it is precisely the dual perspective of application through the prevention of criminal and violent recidivism and of treatment interventions on the recidivist offender that are two of the central objectives to which national and European criminal justice systems are directed³¹³.

Now, an area that has proved particularly flourishing for the exploitation of the potential of algorithms is that of risk assessment³¹⁴, i.e. the prognostic assessment of a defendant's risk-

³¹³ See G. ZARA, *Tra il probabile e il certo*, 17.

³¹⁴ In US criminal doctrine, it is pointed out that 'risk assessment' involves the use of actuarial and algorithmic systems to make predictions about the probability of future crimes. On this point, among others, M. STEVENSON, *Assessing Risk Assessment in Action*, *Minnesota Law Review*, Vol. 103, No. 58, 2018, 314: «The term "risk

recidivism and social dangerousness. Undoubtedly, it should be noted from the outset that, after concluding a cognitive framework of the same, the starting application panorama that will be described consists in looking at the overseas panorama and thus the current and current application uses of risk assessment in the US justice system. Undoubtedly, the US legal system represents a privileged observatory, from which it is deemed appropriate to start the analysis in order to then question the compatibility of automated decision systems with the domestic procedural discipline, as well as with the constitutional guarantees of the individual³¹⁵.

7.1.1 *In an ante delictum application perspective*

When assessing the possible application of these new 'subjects' within the criminal justice system, as anticipated at the beginning of the paper, a twofold perspective opens up from the outset the first aimed at the application of them in the assessment of a subject before the ascertainment of the actual commission of an offence; the second, on the other hand, has regard to all those cases in which the possibility of applying them after the ascertainment of an offence is assessed and therefore in order to support the judge in a more complete assessment of the characteristics of the individual in the phase of commensuration of the penalty in order to ensure a better (more adequate) sanctioning treatment.

First of all, with regard to the possibility of applying such tools in the ante delictum perspective, here, we will limit ourselves to providing a brief overview of a use that is now spreading widely within Italian police headquarters.

It is deemed necessary because it seems appropriate to provide an overview that gathers today the whole or double front on which the debate is focused, especially in the European context.

Well, this premise will therefore be necessary in order to frame from the outset what the possible applicative uses of these instruments, only hinted at here, might be³¹⁶.

This first perspective serves somewhat as a 'big container', as it encloses within it a series of instruments that are being developed within the very locations where crime research takes place and is carried out. On closer inspection, the last few years have seen an increase and stimulation of various instruments aimed at crime research.

assessment," however, usually refers to the use of formal, actuarial, and algorithmic methods of predicting the likelihood of future crime or misconduct».

³¹⁵ P. SEVERINO, *Intelligenza artificiale e diritto penale*, in (ed) U. Ruffolo, *Intelligenza artificiale. Il diritto, i diritti, l'etica*, Milan, 2020, 547 ss.

³¹⁶ The applicability of these in a preventive *ante delictum* perspective will be dealt with in detail in Chapter III.

In fact, the use of AI systems in law enforcement activities is, therefore, already a reality, and indeed it is expected to grow and intensify in the coming years at various levels³¹⁷. After all, the strategic importance of the use of AI systems in law enforcement activities and the valuable results that can be achieved through them are certainly fascinating the 'crime detection' system and the police forces of various countries.

To go into this first perspective, it is undoubtedly necessary to mention what is meant by a new term that has been coined in recent years. 'Predictive policing' can be understood as the set of activities aimed at studying and applying statistical methods with the goal of 'predicting' who may commit a crime, or where and when a crime may be committed, in order to prevent crimes from being committed³¹⁸. Prediction is fundamentally based on an actuarial reworking of different types of data, including those relating to reports of crimes previously committed, the movements and activities of suspects, the locations, the scene of recurrent criminal acts, and the characteristics of these locations, the time of year or the weather conditions most connected to the commission of certain crimes; The data used for these purposes sometimes also include information on ethnic origin, level of schooling, economic conditions, somatic characteristics, which can be traced back to individuals belonging to certain criminological categories (e.g., the age of the victim, the age of the victim's family, etc.), potential terrorists), etc.³¹⁹. In recent times, the use of AI-based software has enabled a quantum leap in predictive policing, since it is now possible to acquire and process an enormous amount of data, uncovering connections that were previously difficult for the human operator to detect³²⁰.

Predictive policing software - whether assisted or not by A.I. systems³²¹- can be basically divided into two categories: those that, inspired by the acquisitions of environmental criminology, identify the so-called 'hotspots', i.e. the places that constitute the possible scenario

³¹⁷ Please refer, in particular, to an interesting study by A. G. FERGUSON, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, New York University Press, 2017, 3 ss.

³¹⁸ See F. BASILE., *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Dir. pen. uomo*, 2020, 17.

³¹⁹ In order to have a more complete overview of the subject of predictive policing, W.L. PERRY-B. MCINNIS-C.C. PRICE-S.C. SMITH-J.S. HOLLYWOOD, *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*, Rand Corporation, 2013.

³²⁰ C. CATH-S. WACHTER-B. MITTELSTADT-M. TADDEO-L. FLORIDI, *Artificial Intelligence and the "Good Society": the US, EU, and UK approach*, 505 ss.; L. BENNET MOSES - J. CHAN, *Algorithmic Prediction in Policing: Assumptions, Evaluation, and Accountability*, in *Policing and Society*, 2016, 1 ss.; G. MASTROBUONI, *Crime is Terribly Revealing: Information Technology and Police Productivity*, 2017; For a concise overview, in Italian, of A.I. systems for predictive policing activities, See R. PELLICCIA, *Polizia predittiva: il futuro della prevenzione criminale?*, in *Cyberlaws*, 9 May 2019.

³²¹ It should be noted that it is not always clear whether, and to what extent, the software we will discuss in the following pages is based on AI systems. This is also due to the fact that some of this software is privately owned and covered by industrial secrecy, so that details on how it works are not made public.

of the future commission of certain crimes; those that, inspired instead by the idea of crime linking, follow the criminal seriality of certain subjects (identified or yet to be identified), to predict where and when they will commit the next crime. It must be said from the outset that, at least for the time being, both systems can only provide adequate predictions in relation to limited, specific categories of crimes (e.g. street crime, such as robbery and drug dealing), and not on a generalised basis for all crimes.

7.1.1.1 The hotspot detection systems

The first type of system includes Risk Terrain Modeling (RTM): an algorithm that, by reprocessing enormous quantities of data on environmental and spatial factors that favour crime, would appear to enable the prediction of the commission of drug offences in certain urban areas. The researchers elaborated this system by submitting to the RTM algorithm data on the environmental and spatial factors most frequently associated with the commission of these crimes: the presence of poor or non-functioning street lights, the proximity of nightclubs, public transport stops, railway stations, high traffic road junctions, ATMs, gold shops, car parks, and schools. This has made it possible to draw up a real 'mapping' of some large metropolitan areas in order to identify the 'hot zones' where the risk of drug dealing is highest, with consequent benefits in terms of planning and implementing interventions to prevent drug-related crime. Similarly aimed at identifying hotspots but in relation to a higher number of offences (not only those of drug dealing) is also a software, already in use for some years in the United States and the United Kingdom, originally developed by researchers at UCLA (University of California, Los Angeles) in collaboration with the local police, and now sold, apparently with great commercial success, by a private American company under the PredPol brand³²².

A device in use by the Italian police would also seem to be inspired by a similar predictive logic: the X-LAW computer system, originally set up by the Naples Police Headquarters, which seems to have already achieved excellent results on Italian territory in the field of preventing certain types of crimes. According to reports, the X-LAW software is based on an algorithm

³²² The tool's website advertises it this way: 'By using only three types of data - type of crime, date/time of crime, and location of crime - to make predictions, PredPol technology has helped law enforcement agencies dramatically reduce crime rates in jurisdictions of all types and sizes, in the US and abroad. PredPol has a proven track record: the Los Angeles Police Department experienced a 20 per cent drop in predicted crime year-on-year, and a local police division was able to experience, for the first time, an entire day without receiving a crime report. The Jefferson County Sheriff's Department reported a 24% reduction in robberies and a 13% reduction in burglaries. In Plainfield, New Jersey, there has been a 54 per cent reduction in robberies and a 69 per cent reduction in car thefts since using PredPol'.

capable of re-processing a huge amount of data extrapolated from the complaints forwarded to the State Police. This reprocessing allows recurring or overlapping factors to emerge, such as the repeated commission of robberies in the same locations, by persons wearing the same type of helmet or motorbike, and with similar modalities. This makes it possible to draw a map of the territory where the highest risk areas are highlighted, up to a maximum level at certain times, thus enabling - in 'hot' areas and times - the police to be prepared to prevent the commission of such crimes and to catch the potential perpetrators in the act.

Indeed, it is necessary to note from the outset, how the *punctum dolens* is to be found precisely in the anxieties of teleological contamination with respect to the constitutional and conventional principles on the use, in investigations, of advanced technical instruments such as satellite tracking, for example. It is precisely in this field that the complex system of balancing the free exercise of personal privacy gives way to the reason of State and, consequently, the lack of legislative criteria, on the modalities of execution, fuels indiscriminate acquisition procedures, since no criteria respecting the rules of use in the light of the criterion of proportionality are provided for.

7.1.1.2 Investigative systems using crime linking

Instead, the Keycrime software, originally developed at the Questura di Milano, and then owned by a private company, is based on the idea of crime linking, following the criminal serialities of certain subjects (identified or yet to be identified), to predict where and when they will commit the next crime.

Other software similarly inspired by the idea of crime linking, and thus of identifying people rather than hot spots, has been developed, and is in use, in Germany (Precobs), in England (Hart - Harm Assessment Risk Tool)³²³, and in the United States. These software tools are based on the basic idea that certain forms of crime are manifested in a very limited time span and geographical area (so-called near repeat crimes): for example, the commission of a robbery would seem to be associated with a high risk of a new robbery being committed, by the same perpetrators and in a geographical area very close to the place of the first crime, within the next 48 hours and, albeit with a decreasing risk rate, up to the whole of the following month. By

³²³ On the HART software, see M. OSWALD - J. GRACE - S. URWIN - G. BARNES, *Algorithmic risk assessment policing models: lessons from the Durham HART model and "Experimental" proportionality*, in *Information & Communications Technology Law*, 2018, 223 ss.; in the Italian doctrine, see M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, in *Diritto penale contemporaneo*, 29th May 2019, 10 ss. The HART software has undergone validation studies by researchers at Cambridge University: see this web address.

collecting and cross-referencing a large amount of data from various sources (e.g. images taken by a camera or information on previous similar crimes), these software tools try to 'profile' the possible perpetrator and predict his next move. Moreover, the results provided by these softwares could in some cases be used not only for predictive purposes, but also to reconstruct the criminal career of the profiled individual, i.e. to have an investigative trail to follow in order to charge him/her not only with the last crime committed (on the occasion of which he/she was identified), but also with the previous crimes constituting the criminal series reconstructed thanks to the storage and processing of data.

7.1.1.3 Brief reflections on the use of such tools for predictive policing

On closer inspection, the predictive policing systems briefly described above can undoubtedly bring great benefits in the prevention of at least some types of crime, but their use raises more than one perplexity. First of all, in fact, it should be noted that their use does not seem to have been regulated so far, in any country, at a regulatory level, so that the conditions and modalities of their use, as well as the evaluation and valorisation of their results end up being entrusted only to practice, and thus to the initiative, sensitivity, and experience of police officers. Yet, their use could entail serious frictions at least with the protection of privacy (in view of the large amount of personal data collected), and with the prohibition of discrimination (to the extent that, for instance, they identify dangerousness factors linked to certain ethnic, or religious or social characteristics)³²⁴. Moreover, these are systems that to a certain extent feed themselves with the data produced by their own use, with the risk of triggering vicious circles: If, for example, predictive software identifies a certain 'hot zone', police checks and patrols in that area will intensify, with the inevitable consequent increase in the rate of crimes detected by the police in that area, which will then become even more 'hot', while other areas, originally not included in the 'hot zones', and therefore not manned by the police, risk remaining, or becoming, for years free zones for the commission of crimes. Moreover, these systems call for crime prevention through active police intervention, through, therefore, a kind of 'militarisation' in the surveillance of certain areas or certain subjects, without, on the other hand, aiming at crime reduction through an action aimed, upstream, at the criminogenic factors (social, environmental, individual, economic, etc.).

³²⁴ On these aspects, A. BONFANTI, *Big data e polizia predittiva: riflessioni in tema di protezione del diritto alla privacy e dei dati personali*, in *MediaLaws*, 24th October 2018; E. THOMAS, *Why Oakland Police Turned Down Predictive Policing*, in *Vice.com*, 28th December 2016; J. KREMER, *The end of freedom in public places? Privacy problems arising from surveillance of the European public space*, 2017.

8 *Brief notes on post delictum application uses (Segue)*³²⁵

As regards the second application guideline, it is only premised here that all those instruments that are applied at the sentencing phase are included. In particular, this includes all those tools that are applied at the pre-trial stage with regard to the decision on the granting of personal liberty or the application of security measures. In addition, the algorithms and A.I. tools on which one can focus most attention concern those that are applied on the choice of punishment.

8.1 An overseas perspective: current applications of risk assessment tools in sentencing and recidivism risk assessment

It should be noted from the outset that predictive justice systems are currently applied in various legal systems. In the investigation carried out in this paper, we shall dwell, in particular, on the comparison between the US model and the European model, attempting to analyse and evaluate in a comparative key the different criminal policy choice adopted. Undoubtedly, the basic premise is given by the recognition of two realities that are remarkably distant in terms of the dissemination and use of technologies in the criminal sphere, and not only. In fact, in the United States, the admission and subsequent application of predictive tools is widespread³²⁶, albeit within certain limits; in Europe, on the other hand, there is a much more cautious attitude with regard to the practicability of these tools, particularly with regard to individual guarantees. Common in both models, however, remains the attitude (still to this day) of scepticism towards the possible application or introduction of such instruments in the delicate phase of ascertaining the antea of criminal liability.

8.1.1 The US model: between evidence-based practice and systems used

As already mentioned, there are several systems based on artificial intelligence mechanisms being used in the United States³²⁷. In particular, the use of predictive justice systems in the

³²⁵ Refer to Chapter 3.

³²⁶ Suffice it to say that it is a trend that the United States has moved towards in recent years in urging the introduction and application of such instruments. The Financial Model Penal Code of the American Law Institute, revised in 2017, did not miss the point, urging the use of 'actuarial instruments or processes to identify offenders who present an unusually low risk to public safety'. Refer to the Model Penal Code, Proposed Final Draft, 10th April 2017, 171.

³²⁷ In recent years, there has been a veritable explosion in the use of algorithms in American criminal justice¹⁴. To realise this, one only has to think that between 2012 and 2015, 20 laws in as many as 14 states 'created or regulated the use of risk assessments during the pretrial process'. For their part, a number of very important associations - including the American Bar Association, the National Association of Counties, the Conference of State Court Administrators, and the Conference of Chief Justices - have spoken out in favour of the use of such tools in the pre-trial phase. Thus, on the subject see the recent article by A.Z. HUQ, *Racial Equity in Algorithmic Criminal Justice*, in *Duke Law Journal*, 2019, 1043 ss.

criminal sphere has already been applied for some time, although limited to the following sectors: pre-crime, in the management of public safety, which therefore includes all the tools of so-called predictive policing; the pre-trial, which includes within it all those tools used to issue an assessment of the prognosis of dangerousness functional to the application of custodial precautionary measures (the so-called pre-trial decisions); and, finally, in the delicate phase of commensuration of the sentence referred to the judge (which includes, per se, also the assessments on the risk of recidivism).

In particular, it is already anticipated here that risk assessment mechanisms are now applied in all phases of the North American criminal trial, whenever a predictive judgement must be made: from the assessments on the release of the defendant, to the sentencing phase, to the judgement on the application of parole or other forms of probation³²⁸.

In the first place, the so-called predictive policing tools³²⁹ come to the fore, which refer to that whole set of techniques and methods, mainly based on statistical operation, that are used by public security authorities to prevent the commission of crimes. In this particular field, which will not be the subject of a specific analysis in this paper, prognosis is based on the interaction of data, including those relating in particular to the possible background, movements and activity of suspects³³⁰. In particular, the use of such software has been particularly welcomed, especially in recent years, in the awareness of the possible advantages that such tools can actually bring in terms of security and prevention.

Secondly, pre-trials come to the fore. In this case, EISs are used to predict whether or not the defendant will refrain from committing new crimes during the proceedings; this type of assessment is undoubtedly relevant to the possible application of personal precautionary measures and to the possible release on bail³³¹.

³²⁸ V. B.L. GARRETT – J. MONAHAN, *Judging Risk*, in *California Law Review*, *Forthcoming*, 9.

³²⁹ *Segue*, 7.1.1 ss.

³³⁰ Così sul punto, F. BASILE, *Intelligenza artificiale e diritto penale*, 1 ss; ID., *Diritto penale e intelligenza artificiale*, in *Giur. It.*, 12, 2019, 67 ss. In particular, it should be noted how, in recent years, algorithmic risk assessment tools have been applied by the US public administration, for example, to indicate to the police in real time the possible risks of committing a crime, according to and applying probabilistic criteria, the areas of the cities to be controlled or garrisoned (the so-called crime mapping). Other predictive policing tools also include the so-called 'no-fly list', i.e. an application that collects and analyses data on potential terrorists, to prevent possible attacks. On this topic also, R. FLOR, *Le nuove frontiere del contrasto alla criminalità: dalle investigazioni tecnologiche alla predictive policing al servizio della Urban Security*, in T. Dalla Massara-M. Beghini (eds), *La città come bene comune*, Naples, 2019, 179 ss.

³³¹ In particular, the tool among these that has found the greatest application to date is the PSA. See PERRONE D., *La prognosi postuma tra distorsioni cognitive e software predittivi. Limiti e possibilità del ricorso alla "giustizia digitale integrata" in sede di accertamento della colpa*, Turin, 2022, 84.

Lastly, predictive algorithms have also taken on a central role in the extremely delicate sentencing phase, i.e. the phase in which the judge must decide on the quantum of the sentence. On this point, just to anticipate what will be said in the following chapter, it should be noted that towards the end of the 1990s, in order to prevent the frequent biases into which popular juries 'fell', the conviction became widespread that decisions on the treatment of punishments, on the granting of possible rewards and on alternative measures to detention, should be based on statistical evidence. Indeed, the use of such technologies has been subject to various criticisms arising from their first applications.

Indeed, it appears useful to focus on a model, or rather an approach, that has already been mentioned based on so-called evidence-based practices.

On closer inspection, the evidence-based assessment aims at ascertaining the criminal dangerousness of a certain subject; in fact, it presupposes as a first necessary element the identification of a series of factors (or also called predictors) directly involved in criminal behaviour, which may concern several elements age, gender, ethnic origin, level of schooling, work and family situation, economic (or income) level, criminal record, places and people frequented, the presence of offenders within the restricted family nucleus or in the network of acquaintances, place of residence, any history of violence and other contextual variables continuously subject to variability (such as, for instance, lack of family and social support), drug or alcohol consumption. The predictors mentioned are clearly not univocal and do not always maintain the same variability or static nature; they may in fact present different rates of dynamism, in the sense that there are static (and non-modifiable) factors such as gender and ethnic origin, for example, and other dynamic factors that are modifiable and vary over time (such as the neighbourhood in which one lives or the places and people one frequents). Finally, there are other types of risk factors, also referred to as acute, that change rapidly over time and are associated with a condition that facilitates a violent reaction (e.g. drug use).

These factors, once they are collected and statistically processed and weighted, are, according to this approach, combined following a so-called actuarial approach in order to obtain a certain score that is associated with a different 'scale' that attributes an indicator of dangerousness to the subject being examined³³².

Interestingly, this peculiar type of approach was borrowed from the insurance industry (based precisely on the quantification of different types of risks); it is not a new approach for

³³² On this point, G. ZARA, *Tra il probabile e il certo. La valutazione dei rischi di violenza e di recidiva criminale*, DPC, 20th May, 2016.

the American legal system; Indeed, up until the 1920s, the US criminal justice system used various factors such as age, race, criminal history, occupation (even the school grades of individuals) to predict which former inmates had or possessed certain characteristics indicative or symptoms of being more dangerous than others, or even to determine whether they required certain clinical treatment (e.g., drug use) upon release.

Today, on the other hand, these actuarial assessments in most US states are carried out using an actuarial approach³³³ and, before that, the collection and processing of the data that allow the risk scales to be prepared and are therefore entrusted to artificial intelligence systems and in particular, as already mentioned, to predictive algorithms³³⁴.

However, it should also be noted that there is a very heterogeneous multiplicity of predictive algorithms in American jurisdictions; they differ from each other in that they take into account different risk factors (in some cases more static ones, in others dynamic ones)³³⁵.

For systematic reasons, on the other hand, only those risk assessment systems that are most widely used or can serve as a useful model for the analysis to be performed here will be considered.

8.1.1.1 California's position: the inauspicious outcome of Proposition 25 on replacing the cash bail with pretrial risk assessment tools

On closer inspection, although the panorama presented would seem to see the US as the great proponents of these applications, however, the reality is not always so homogeneous and some issues have already arisen in some US states.

To put a brake, albeit a minimal one, on this attitude of great fervour towards predictive tools, California's position of 3 November 2020 appears peculiar, offering in part a position of diffidence towards such tools on the part of states that have already been applying them for some time.

³³³ Because these models are generated on the basis of statistical correlations, not causal links: if an algorithm detects, for example, that low income correlates with high recidivism, it does not mean that being poor causes criminal behaviour; instead, this is exactly what risk assessment tools do: they transform correlative intuitions into causal scoring mechanisms. As a result, groups that have historically been disproportionately targeted by law enforcement - particularly low-income and minority communities (blacks, Hispanics, etc.) - are at risk of being penalised by disproportionately high risk scores, and thus end up/stay in jail more than other social groups.

³³⁴ See also on this point, R. WERTH, *Risk and punishment: The recent history and uncertain future of actuarial, algorithmic, and "evidence-based" penal techniques*, 10th January, 2019.

³³⁵ It was calculated, for example, that in 2015 more than 60 different risk assessment tools were applied for the sentencing phase alone. A.Z. HUQ, *Racial Equity*, 1075.

In fact, with Proposition 25, the choice on the repeal of the cash bail³³⁶ law was put back to the citizens of California. In essence, what was put back to choose was the permanence of a predictive algorithm capable of calculating the degree or different level of risk of the bailout.

The idea at the basis of the reform was to try to restore, in a certain sense, conditions of substantial equality: in fact, for the promoters of the reform, deciding to subordinate the maintenance of a measure of pre-trial detention to the economic impossibility of the defendant to support the economic commitment of the bail meant precisely to favour the permanence of discriminatory situations to the detriment of the less well-off; in this way, the rationale of pre-trial detention of the dangerousness of the defendant or the danger of flight was unjustifiably undermined³³⁷.

Therefore, one also notes how the legislative text referred, in particular, to the use of risk assessments corroborated and validated by scientific research which would then have issued a given result placing individuals on different levels of risk of recidivism and dangerousness: low, medium and high. What is noticeable is that the most significant fact concerns the role that these systems would have played in the judicial determination; indeed, it was not a question of replacing the judicial body, but the result then generated by the same algorithm would have been reviewed by the judge; the judge would still have had wide discretion in the assessment underlying the final decision³³⁸.

³³⁶ The latter provided for the defendant to be held on bail pending trial, and in its place the California Money Bail Reform Act, also known as Senate Bill 10 (SB 10), which would have made the continuation of pre-trial detention conditional on the existence of a flight risk calculated by algorithmic risk assessment systems; if the referendum was successful, California would have become the first state in America to have amended this legislation by replacing cash bail with the use of algorithmic risk assessment systems. Thus, T. Fuller, California's the First State to Scrap Cash Bail, in *New York Times*, 28 August 2018. It is worth noting, however, that in many California jurisdictions the cash bail system is accompanied by the use of risk assessment tools. For this aspect and an analysis of the dreaded dangers of the reform, see T.A. MERKL - L. ARZY, *California's Referendum to Eliminate Cash Bail, Explained*, 2nd October 2020. The document predates the referendum and in particular highlights the recommendation not to use risk assessment tools as an alternative to cash bail in the event of an outcome against the amendment of the bail law and thus future reforms, in order not to foster racial discrimination that would underlie the calculations made by such systems.

³³⁷ In addition, it should be noted that the novelty developed in parallel with the jurisprudential evolution that progressively manifested itself in the same direction in the Californian criminal courts, inaugurated by the judgment of the San Francisco Court of Appeal *Humphrey on Habeas Corpus* which, recalling the legislative direction already expressed several decades ago on the need for reform, upheld the defendant's appeal affirming the unconstitutionality of the bail law for conflict with the corollaries of due process. In addition, Senate Bill 10 provided for the establishment of pretrial assessment bodies in the Superior Courts ('Pretrial Assessment Services'), which would be entrusted with the task of assessing the risk of recidivism or the danger of flight³⁰ and making recommendations for conditions of release. On the definition of "risk"; see Senate Bill 10 (SB 10), Article 1. Definitions, § 1320.7 (h): "'Risk' refers to the likelihood that a person will not appear in court as required or the likelihood that a person will commit a new crime if the person is released before adjudication of his or her current criminal offense".

³³⁸ A.S. NIEDERMAN (et oths)., *The Institutional Life of Algorithmic Risk Assessment*, in *Berkeley Technology Law Journal*, Vol. 34, No. 3, 2019, 719; However, the authors point out that the design and use of risk assessment tools nevertheless restrict the discretion of the adjudicating body because they involve technical and political choices.

What is relevant about this situation is the outcome of the referendum, as the majority rejected the introduction of this risk assessment³³⁹.

What is certainly relevant is that in a balancing of the interests and rights at stake, a pretrial law that was already in itself going to generate discrimination because it was directly dependent on the economic situations of the subjects, the situation of a country that refuses to introduce a Reform fearing, even in that case, further discriminatory drifts arises even more strongly and markedly.

9 *Other risk assessments used in the investigation phase*

It is only necessary to provide a brief introduction to a part of the evolution that has characterised risk assessment tools.

Indeed, as already mentioned in the descriptive part on risk assessments and their evolution, alongside and in parallel with the long and complex evolution of these instruments, more recently risk assessment tools have also been developed and applied for use in the investigation phase (in particular for bail decisions).

However, interim forms of deprivation of liberty are applied in order to prevent 'procedural risks', such as the risk of absconding and evidential pollution. The underlying idea is always centred on the possibility of predicting the future behaviour of the subject in a given court decision. In such a case, it is not only a question of assessing the dangerousness of the individual himself, but also what the possible damage might be if he were left at large. This is certainly a different type of risk assessment from the one mentioned above.

9.1 *An example of risk assessment: the Public Safety Assessment (PSA)*

Public Safety Assessment (P.S.A.), is an algorithmic social dangerousness assessment system³⁴⁰. The PSA is the most widely used instrument for determining the application of pre-trial supervision measures and release on bail.

³³⁹ The result was undoubtedly rather strange, since the referendum ended with a significant prevalence of votes against the reform, 56.41% of the voters against 43.59%, even though the use of pretrial risk assessment tools is recurrent on a large scale in the California Criminal Courts, as shown by a December 2019 report by the Public Policy Institute of California, according to which 49 counties out of a total of 58 were already using pretrial risk assessment tools.

³⁴⁰ This tool was devised by the Laura and John Arnold Foundation, a non-profit organisation. It is currently used in three states and twenty-eight jurisdictions. Thus on the point, D. KEHL – P. GUO – S. KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing, Responsive Communities Initiative*, 10.

This tool assists the judge in deciding whether to release the accused before the trial is defined and concluded. The objective for which it was created and easily applied was to reduce the number of pre-trial detainees.

It is based on an approach that can be either actuarial or non-actuarial and that "compares the risk factors of the subject undertrial with a database of 1.5 million cases from three hundred jurisdictions across the United States and, based on the information available, scores the subject on a scale of one to six. There are nine risk-measuring factors examined (including age, criminal record, past court appearances, and complaints received in previous cases), and neither race nor ethnic or geographical origin appear among them³⁴¹.

Most famous for being the system by which the State of New Jersey reformed its parole, using it as an alternative to bail³⁴². It acts as an aid to the judge and would have led to an increase in the number of releases on parole without bail³⁴³. In fact, the first state to adopt the P.S.A.³⁴⁴, in the pre-trial phase, was Kentucky in 2013³⁴⁵.

The peculiarity of this tool is that it is based on only nine factors; it is an actuarial tool and the factors it takes into account are: the individual's age, pending charge, and criminal history. The peculiarity lies in the fact that elements such as place of birth, ethnicity are not taken into account as it was felt that they could be detrimental to the accuracy of the prediction. Furthermore, this tool does not require an interview with the accused, as information can be extracted on objective data, such as pending charges. It serves to predict failure to appear in

³⁴¹ F. BASILE., *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *DPU*, 2020, 17.

³⁴² On this point, E. LIVNI, *Nei tribunali del New Jersey è un algoritmo a decidere chi esce su cauzione*, in *Internazionale*, March 2017 (trad. F. Ferrone) available at the following link: <https://www.internazionale.it/notizie/ephraat-livni/2017/03/03/tribunali-algoritmo-cauzione>.

³⁴³ F. BASILE., *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *DPU*, 2020, 19.

³⁴⁴ In this text, the factors used and the weight of each factor are explained. S. QUATTROCOLO, *Artificial Intelligence, Computational Modelling and Criminal Proceedings. A Framework for a European Legal Discussion*, Berlin, 2020, 151: "The risk factors considered by PSA under this parameter are: Pending charge at the time of offense (Y/N); Prior conviction (Y/N); Prior failure to appear in the past 2 years (No/Once/2 or more times); Prior failure to appear, older than 2 years (Y/N)". The use of this software, developed in 2013, was supposed to be helpful in reducing the number of individuals detained before trial, but the results are considered to be poor "the movement for bail reform is one rare area of bi-partisan agreement, pushing for abandoning monetary bail, in favour of effective management of risk"; C FAZEL, *The scientific validity of current approaches to violence and criminal risk assessment*, in *De Keijser*, Roberts, Ryberg (eds), *Predictive sentencing, normative and empirical perspective*, Oxford, 2019, 197, which shows that 39 federal states have their own 'risk assessment tool'; in England and Wales, OASys is the offender assessment system routinely used in the National Offender Management Service (NOMS), within which different risk assessment tools are used, depending on the criminogenic needs of the offender assessed. G. CONTISSA – G. LASAGNI – G. SARTOR, *Quando a decidere in materia penale sono (anche) algoritmi e IA: alla ricerca di un rimedio effettivo*, in *Riv. trim. diritto di internet*, No. 4, 2019.

³⁴⁵ Developed by Arnold Ventures on the basis of the largest and most diverse set of pre-trial records ever collected (approximately 750,000 cases from around 300 US jurisdictions) and validated using over 500,000 cases from multiple jurisdictions, the Public Safety Assessment (PSA).

court pretrial, new criminal arrest while on pretrial release, and new violent criminal arrest while on pretrial release.

Virginia has also been at the forefront of the use of risk assessment at this stage of the trial process and requires that the results must be checked by public officials before they are published, thus adopting a structured professional approach³⁴⁶.

Specifically, the predictive system considers three possible outcomes:

- The individual's failure to appear: based on charges pending at the time of arrest, previous convictions, failure to appear in the past two years and more;
- The individual's new criminal activity: based on charges pending at the time of arrest, previous convictions for violent and non-violent offences, failures to appear in the past two years, previous prison sentences, whether the individual was young at the time of arrest;
- new 'violent' criminal activity: based on charges pending at the time of arrest, previous convictions for violent and non-violent offences, violence of the offence charged and possible young age at the time of arrest.

Therefore, a further peculiarity of these instruments is that they produce a risk assessment scale based on three different parameters:

- FTA: Failure to Appear
- NCA: New Criminal Activity
- NVCA: New Violent Criminal Activity

This scale operates on several levels and can be used by the judge together with another tool, the so-called Decision Framework, to decide more comprehensively whether the arrestee will be released or detained³⁴⁷.

Each of these possible outcomes is given a score from 1 to 6, except for the third outcome which results in a yes/no; the scores combined then give a total score, resulting in a specific recommendation for each defendant³⁴⁸.

³⁴⁶ M. DEMICHELE (et oths)., *The Public Safety Assessment: A Re-Vaulation And Assessment Of Predictive Utility And Differential Prediction By Race And Gender In Kentucky* 48, 2018, 17.

³⁴⁷ Apart from any consideration of the apparent violation of the presumption of innocence inherent in the NCA, NVCA indices, the particularity of this instrument, compared to those considered so far, is its alleged ability to provide a 'Failure To Appear' index, on this point S. QUATTROCCOLO, *Sui rapporti tra pena, prevenzione del reato e prova nell'era dei modelli computazionali psico-criminologici*, 271.

³⁴⁸ *Ibidem*, 18.

This software is free of charge and was designed from the outset as a support for the criminal judge for pretrial decisions, particularly those concerning bail or pre-trial detention³⁴⁹.

P.S.A. was created primarily as a response to the criticism and doubts that predictive algorithmic tools have attracted in the United States since their appearance; in fact, it was created with the intention of reducing pre-trial detention and to provide an alternative to the practice of bail.

Its dataset does not include data on gender, race or social class, and both the source code and the data processing procedure are public.

A small percentage of bias is present in the predictive model³⁵⁰, but it seems to be in reasonable terms, and this was probably influenced by the fact that the developer is a non-profit organisation.

In fact, this tool is one of the most positively criticised at the moment; indeed, in Lucas Conty, Ohio, where the tool was adopted in 2015, a significant increase in the number of people set free, without recourse to bail, and a decrease in the number of offences committed while awaiting trial were noted³⁵¹.

What has certainly been noted is that the great success and spread that this instrument has achieved in the United States is certainly, if only partially, connected to the profound crisis that the bail institution is going through, which has shown the first signs of weakness due to its serious discriminatory effects. This crisis³⁵² has created fertile ground for the spread of such instruments in an attempt to make certain types of assessment more 'objective'.

³⁴⁹ On closer inspection, these decisions in the US have enormous consequences both for the individual accused of a crime and for the community at large, as spending only a few days in prison can cost jobs, housing and health services and significantly disrupt family life. Moreover, studies show that people detained before trial are more likely to plead guilty, be convicted and be re-arrested (9). Since the US Supreme Court has ruled that pre-trial liberty is the norm and detention should be the carefully limited exception, the company that created the software felt that "the key factor to consider when making these pre-trial decisions is the likelihood that the person will not flee the jurisdiction and/or pose a danger to others" and, far from minor, that "a person's inability to post bail should not determine whether they remain in c Since its development in 2013, PSA has been implemented in dozens of jurisdictions across the country including the states of Arizona, Kentucky and New Jersey, and in some of the largest cities such as Phoenix, Chicago and Houston.

³⁵⁰ *Ivi* 54, 56.

³⁵¹ However, there is no shortage of critical points to be reported here as well. In this sense, again, J. TASHEA, *Risk-Assessment Algorithms challenges in bail, sentencing and parole decision*, March, 1, 2017, who reaffirms that «even if an algorithm is equally accurate for all, more blacks and males will be classified as high risk because African-Americans and men are more likely to be arrested for a violent crime».

³⁵² The evaluation of pre-trial risk assessment tools is based on elements that are largely different from those taken into account for traditional risk assessment tools. First, referring to the most popular pre-trial risk assessment tool in the United States at the time, the PSA, it is worth noting that its main difference from other actuarial tools is the FTA index. However, the factors taken into account in assessing an individual's propensity to flee are not supported in the literature by any empirical evidence of actual relevance in terms of the risk of the defendant failing to appear at the hearing. Whereas the risk factors of criminal behaviour, together with protective factors, criminogenic needs, and correlates, have been the subject of wide-ranging and in-depth empirical scientific research for about a century, there seems to be no evidence - in the scientific literature - that convincingly

Clearly, it is very difficult to establish parameters that are capable of assessing the goodness and correctness of the choices made by the judge with regard to pre-trial risk since there are very few, in most cases, data on which he can base himself³⁵³.

9.2 *The PATTERN algorithm: the system to be taken as a model?*

On closer inspection, a new system that has found application in the USA is the PATTERN system (Prisoner assessment tool targeting estimated risk and needs). The same, in fact, represents an algorithmic risk-recidivism assessment tool of an institutional nature and for this reason, therefore, not covered by industrial secrecy³⁵⁴. The peculiar element that is presented in this tool and that would seem, for the moment, to give it an added advantage over the others in use is given by the fact that when it was created, thought was immediately given to the situation and problems that were generated with such tools, first and foremost the problem of opacity and transparency. On this point, in fact, the DOJ would seem to have tried to overcome this problem. This would result, albeit in part, from these elements: during the development of the tool, three listening sessions were held with the aim of listening to and interacting with experts, stakeholders, including organisations representing victims of crime³⁵⁵.

demonstrates the relevance of FTA factors. Evidence is apparently lacking to prove that the incorporation of these factors into an actuarial tool is capable of outperforming the judge's individual assessment in terms of reliability. Moreover, such a finding is particularly difficult to obtain. The comparison between the results of the PSA and the judicial decisions taken by individual judges is vitiated by the fact that the failure to obtain a bail may depend either on the judge's assessment, more or less correct, of a high risk of absconding, or, more banally, on the overestimation of the economic capacity of the arrested person, on whom a condition has been imposed that he is unable to fulfil. In addition, the judge's individual assessment may be influenced by factors other than risk, such as proportionality to the pending charge, which, in the face of a risk of absconding or criminal behaviour that is not indifferent, may not justify the pre-trial detention of the accused.

³⁵³ Therefore, due to these arguments, some scholars have questioned whether it is possible to estimate whether pre-trial risk assessment tools really outperform humans in predicting endo-procedural risks. Moreover, since tools such as the PSA rely exclusively on information extracted from archives and records, without the need for an interview with the defendant, some authors have criticised its structure. It would seem contradictory, in fact, to claim the overcoming of the pecuniary bail system by suggesting instruments that are completely based on data extracted from the very system that is the subject of the reform itself.

³⁵⁴ Basically, most of the instruments used in the US are covered by trade secrets. On this point, see among others, R. WEXLER, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, in *Stanford Law Review*, Vol. 70, No. 5, 2018, 1343 ss; A. RIZER - C. WATNEY, *Artificial Intelligence Can Make Our Jail System More Efficient, Equitable, and Just*, in *Texas Review of Law & Politics*, Vol. 23, No. 1, 2019, 214- 215. Please also refer to the statement by D.E. PATTON, *Executive Director del Federal Defenders of New York, Oversight Hearing on "The Federal Bureau of Prisons and Implementation of The First Step Act"*, 2019, 3,8, who observed that the secrecy of black box models casts doubts on the reasonableness of the results produced by the algorithm: «Across risk assessments in criminal justice, the secrecy that permeates black box instruments causes significant concerns about how reasonable they are in practice».

³⁵⁵ In the report published on 15 January 2020, the DOJ thus announced that the updates and improvements made to PATTERN were the direct result of the suggestions and solicitations from those bodies involved in the process of developing the new system. On this point, see 5 See the DOJ report, *The First Step Act of 2018: Risk and Needs Assessment System*, 2019, and the press release of 15 January 2020, *Department of Justice Announces Enhancements to the Risk Assessment System and Updates on First Step Act Implementation*.

On closer inspection, although it may not be possible to make predictions now, it can be seen from the outset that PATTERN in fact represents a significant step forward towards overcoming one of the major and seemingly insuperable criticalities that had been raised by the systems used in the various stages of criminal proceedings, namely the very lack of transparency. Undoubtedly, this instrument is part of a precise legislative and political line of trying to overcome the major criticalities already presented in other contexts by such instruments and to make (it is believed) this system more correct and efficient³⁵⁶.

What certainly appears to characterise the overseas system a great deal is how much trust Congress itself places in algorithmic evaluations and results within a penal system³⁵⁷.

10 *Brief remarks: the anticipation of an initial operating proposal (Segue)*³⁵⁸

It seems opportune to premise that, after a review of the historical and scientific evolution of risk assessments, it is immediately apparent that these, although developed in the United States, were the end product of psycho-criminological theories and studies on incapacitation.

Indeed, although there is a tendency to link the risk of recidivism with incapacitation, in reality, if we want to return to their original rationale and use of risk assessment, the real purpose of 'risk assessment' should be to be able to guide a specific intervention on the subject in order to implement the prevention of future antisocial behaviour, indicating to the same subject alternatives that are attractive with respect to crime³⁵⁹. Obviously, if the risk assessment is the more reliable, the more this objective can be achieved.

The starting point for starting and improving the reliability and validity of studies on the risk of violence and criminal relapse sees and has seen the development of four steps. Firstly, in fact:

1. Studying a wide range of different risk factors (continued in section 5.1.)

³⁵⁶ Remaining on the topic of reforms, this time on a federal level, in the United States great expectations are now being placed on the impressive criminal justice reform implemented with the First Step Act (Formerly Incarcerated Reenter Society Transformed Safely Transitioning Every Person Act) of 21 December 2018, which aims to counteract the phenomenon of mass incarceration and purge it of distortions stemming from racial discrimination; The project is ambitious in light of its objectives, which include limiting the use of restrictive measures by favouring reintegration programmes developed by the Department of Justice. The First Step Act has entrusted algorithmic systems with a central and indeed decisive role in achieving its aims: in Title I, Section 101, with § 3632 of the enactment, entitled Development of risk and needs assessment system, Congress authorised the Attorney General of the United States to develop a new risk assessment system for use by the Bureau of Prisons (BOP), instrumental to early release and to grant sentence reductions to individuals detained in federal correctional institutions who present a 'low' or 'minimal' risk of re-offending.

³⁵⁷ They are also defined as 'objective and statistically validated' instruments.

³⁵⁸ Refer to Chapters 4 and 5.

³⁵⁹ G. ZARA – D. FARRINGTON, *Criminal recidivism: explanation prediction and prevention*, 151.

2. Broaden the criterion for measuring relapse or criminal relapse, which should consist of more than the variable 'new arrest' or 'new conviction', also including self-report behaviour, information on family history, medical, psychiatric and hospital records;
3. Involve both male and female participants; acquire different data from multiple sources.

This, without a doubt, represents the starting point and the assumption on which the subsequent scientific research was built, which made it possible to overcome the enormous number of false positives that had been found in many of the unstructured assessments of the risk of violence and the risk of criminal relapse.

A first result is to consider the baseline rate of violence in general and in the target population. In particular, it should be specified that the base (or prevalence) rate of a particular behaviour, in a particular population, means the proportion of people who - at a particular and given time - manifest that behaviour. It is all the more necessary to consider the base rate as it provides a measure of occurrence without having a comparative value³⁶⁰.

A second result that has been achieved through structured utilisation is a growing understanding of the nature and risk processes of persistent violence and their temporisation. In fact, persistent violent behaviour is the result of a systematisation over time of an externalised pattern of aggression, hostility, impulsiveness, destructiveness, which becomes much more unlikely in psychiatrically treated and pharmacologically followed persons. As has been medically proven, the risk of violent relapse in psychiatric patients is high in the period immediately following the violent act (generally between 24 and 48 hours). Indeed, some scholars suggest that only a very small portion of violence acted out by mentally ill individuals is directly caused by the symptoms of the illness: only 10% of cases.

Another and final result concerns the specificity or otherwise of the predictors of violent behaviour in the psychiatric population: clinical-forensic studies have shown that people with mental disorders share the same risk factors and processes as the mentally healthy population. Indeed, it did not take long to convince the researchers that special attention should be paid to those risk factors that are directly involved in violent behaviour and that related to gender, age, career or criminal history, difficulty in regulating anger and aggression, impulse control, early onset of conduct disorder, comorbidity with drug behaviour, a history of previous acted out

³⁶⁰ This means, to give a practical example that if, for example, the rate of violence in the general population is 2%, this means that, on average, 2 out of 100 people will act violently. In this case, calculating the basic rate of violence in the psychiatric population only makes sense when compared to the general population. In fact, the probability of a single event leaves the reference class indeterminate by definition; a reference, on the other hand, is always necessary to interpret or make a probability estimate with respect to the need to understand how likely it is that the prisoner (or patient) who possesses certain characteristics will commit an act of violence in the next six months, if, for example, he were to be granted a weekend home leave.

violence, history of hospitalisation, pro-criminal thinking, and certain contextual variables such as lack of family and social support. In addition, substance use, non-compliance with treatment and psychopathy were also found to be particularly significant predictors of criminal and violent relapse in both psychiatric and non-psychiatric individuals.

A final result was to recognise the need to construct and test diversified instruments for samples of individuals differing in age and gender, in the type of criminogenic needs present, in the different assessment, treatment and therapeutic contexts, and for the different purposes involved.

To date, it is necessary to look at the fourth-generation risk assessment tools. In an attempt to provide, albeit at this juncture, only a cursory overview of their structure, it should be noted that, first and foremost, the main focus around which the functioning mechanism of risk assessment revolves is the assessment, management, formulation and reduction of risk and compliance. The purpose towards which its use is directed is the integrated analysis in which the evaluative accuracy is associated with the evaluative specificity of the individual case for the purposes of treatment and prevention.

On the other hand, the historical risk factors that are used and taken into analysis concern the integrated assessment (actuarial and individualised professional) of the historical, clinical, contextual factors, which are recognised as empirically relevant and are, moreover, considered to be significant for the treatment purposes of the specific individual case. In this case, attention is paid not only to the conditional risk, but also to the conditions and specifications influencing the transition of the risk to its actualisation (the so-called risk formulation).

The clinical reference factors, on the other hand, relate to an integrated assessment (actuarial and statistical and structured professional) of the traits, symptoms and resources recognised as empirically relevant and significant in the treatment of the specific case and in the promotion of compliance.

Instead, the factors of relevance (which originally in the first generation were only the extraordinary and unexpected ones) now concern and encompass a much broader panorama; in particular, they concern all criminogenic factors present and the planning of intervention aimed at risk reduction and the promotion of activatable protective factors.

The method put into practice, on the other hand, relates to a novelty with respect to previous uses: in particular, it concerns a structured professional and clinical interview, aimed at the assessment, collection and integration of collateral information (the so-called file review). It also concerns the identification of areas on which to intervene, data integration and inter-professional comparison, intervention planning and response analysis.

10.1 *The latest generation of risk assessment tools: why are they better than others in criminal risk assessment?*

Undoubtedly, the proposal to be presented here has risk assessment tools as its protagonists, and in particular the typology of actuarial judgments. At the outcome of a study that has analysed the various typologies, it is believed that the latter are the most appropriate and advantageous to use in providing a long-term probability estimate of violent, criminal or even hetero- and self-directed aggressive behaviour³⁶¹. Indeed, the actuarial method guarantees the expert to be able to operate a real risk mapping that is based on those static risk factors relevant in the commission of criminal and violent behaviour, precisely collected and measured; it also allows the identification of those individuals most at risk and most in need of monitoring, observation, support and therapy. In fact, it is believed that the structured collection of historical, anamnestic data, linked to the criminal career of the person, can help the expert in a twofold direction on the one hand, in order to avoid both judgmental attitudes and negative counter-transference reactions, not so infrequent with such types of criminal individuals; on the other hand, reactions of unsuitable and excessive benevolence that confuse the professional responsibility of taking charge with a sort of 'passive and paternalistic'³⁶² welfarism that risks depriving the person of responsibility and triggering dependency reactions.

The completion of the assessment will then be all the more accurate and ethically sound the more it is focused on those risk factors that are scientifically and clinically relevant in defining the type of criminal career and in activating a specific treatment plan.

Aspects that are of crucial importance in the transition between the scientific validity of risk assessment and its practical application in criminal justice therefore deserve careful consideration. A first aspect is that the match required is not between a single piece of scientific data and another, but between this and the appropriate 'legal convention'; although the assessment of human behaviour is then based on conventionally and legally rubricated aspects, it can only be conventional, even if enriched by 'scientific evidence'. Indeed, the scientific approach requires generalisations; the clinical approach, on the other hand, requires individualisation, and, finally, psycho-forensic work requires a 'meta-analytical' step. This means that the individual case, in its individuality and singularity, must be examined in the light of a systematic, explicit, context-appropriate, valid, reliable and specific scientific

³⁶¹ This condition is described as the 'status of the criminal individual' by K. S. DOUGLAS – J. L. SKEME, *Violence risk assessment: getting specific about being dynamic*, in *Psychology Public Policy and Law*, 11, September 2005, 347-383.

³⁶² G. ZARA, *Tra il probabile e il certo*, 69.

method. In all cases and steps, however, steps must be taken to minimise the loss of specificity and to protect methodological accuracy as far as possible³⁶³. Mere clinical intuition, which is so important in a treatment setting, would not suffice and would be sufficient, but it would be imprudent and almost arbitrary in a forensic context (with all the implications for the individual's guarantees) if it were not backed up by solid, objective and quantifiable data, in a dimension that does not deal with the ultimate causalities, but instead intends to bring together current knowledge in an integrated evaluative and clinical dimension.

Indeed, it is no coincidence that this so-called actuarial assessment of criminal dangerousness, which constitutes the theoretical prerequisite for the use of predictive algorithms, is becoming increasingly widespread in recent years. It is no coincidence, in fact, that the most widespread concept in recent years is precisely that of evidence-based assessment of the individual risk of committing a new crime; this is because, it is explained, it is based on objective evidence and is probably destined to supplant the intuitive assessments of judges, which have been widely used to date. Indeed, the evidence-based assessment of criminal dangerousness presupposes the prior identification of a series of risk factors (or even predictors) directly involved in criminal behaviour³⁶⁴. The peculiarity concerning such predictive tools is that the factors analysed (e.g. age, gender, criminal record, etc.) once they are collected, thanks to prospective longitudinal studies, can allow an actuarial (or statistical) approach to the assessment of criminal dangerousness. Therefore, precisely through a combination of them, 'scales' can then be set up that allow a certain score to be attributed to the subject under examination.

The element of distinction and merit of these so-called actuarial assessment instruments is the fact that the scales used for the actuarial assessment of criminal dangerousness differ according to the population in relation to which they have been drawn up (e.g. population of adults, of minors, of prisoners, or even according to the time scale of the risk (whether immediate, medium or long-term)³⁶⁵.

³⁶³ The issue of G2i - where G stands for general scientific propositions (framework evidence); i stands for individual (diagnostic evidence) - calls for methodological rigor as the first, indispensable dimension and for clinical and individualised analysis as the *conditio sine qua non* of the transition from evaluation to re-educational intervention and customised treatment.

³⁶⁴ Così, sul punto, F. BASILE., *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, 17.

³⁶⁵ See, G. ZARA, *Tra il probabile e il certo*, 14.

10.2 Overcoming the intelligibility of decision-making mechanisms

The issue of trust in algorithms is assuming increasing importance in the face of the widespread diffusion of these technologies, which are now capable of providing suggestions and directing human action in a variety of ambits³⁶⁶. This is a central aspect for the future development of artificial intelligence, which has also received particular attention at European level with the formulation by the Independent Group of Experts appointed by the European Commission, the Ethical Guidelines for Reliable Artificial Intelligence³⁶⁷. Here, it is strongly emphasised that the issue of trust serves as the de facto pivot in the framework of human rights protection in relation to artificial intelligence. However, as things stand, the road to fully trustworthy systems appears strewn with obstacles. In particular, the debate revolves around several recurring critical issues that are inherent in structural aspects of machine learning-based technologies.

By now a 'classic' within the discussion on artificial intelligence is the argument of opacity, according to which the algorithmic machine is a '*black box*', whose mysterious operation mysterious functioning does not allow one to fully grasp the steps that, starting from a certain input, lead the processing to a certain result³⁶⁸.

This lack of transparency which characterises artificial intelligence - especially deeplearning systems and neural networks³⁶⁹ - leads to a series of problems regarding the function

³⁶⁶ For these aspects, please refer to Y. DUAN - J.S. EDWARDS - Y.K. DWIVEDI, *Artificial intelligence for decision making in the era of Big Data. Evolution, challenges and research agenda*, in *International Journal of Information Management*, 48, 2019, 63-71. In this sense, the experiment conducted is significant da J.M. LOGG - J.A. MINSON - D.A. MOORE, *Algorithm appreciation: People prefer algorithmic to human judgment*, in *Organizational Behavior and Human Decision Processes*, 151, 2019, 90-103, which investigated the propensity of subjects to trust algorithms in different decision-making contexts. In particular, the study focuses on levels of 'algorithm appreciation', assessing whether and when people are willing to rely on the suggestions of an algorithm instead of human judgement. On this point, also T. AARAUJO (et oths), *In AI we trust? Perceptions about automated decision-making by artificial intelligence*, in *AI & Society*, 35, 2020, 611-623.

³⁶⁷ The document is available at <https://op.europa.eu/it/publication-detail/-/publication/d3988569-0434-11ea8c1f-01aa75ed71a1> (last consultation: 06.03.21). The topic has also recently been addressed in the Communication addressed by the European Commission to the European Parliament, the European Council, the European Economic and Social Committee and the European Committee of the Regions, Building Trust in Human Centric Artificial Intelligence, Brussels, 8.4.2019, available at <https://digital-strategy.ec.europa.eu/en/library/communication-buildingtrust-human-centric-artificial-intelligence> last consultation: 08.03.21), where in the foreword it is emphasised that 'trust is a prerequisite to ensure a human-centric approach to AI', again emphasising the link between the implementation of trustworthy artificial intelligence and the EU's goal of protecting human rights.

³⁶⁸ On this subject, the reference to the work of F. PASQUALE, *The black box society. The secret algorithms that control money and information*, Cambridge-London, 2015; the author, with the expression '*black box*', alludes to the unknowability of algorithmic mechanisms, highlighting above all their economic consequences.

³⁶⁹ G. SARTOR - F. LAGIOIA, *Le decisioni algoritmiche tra etica e diritto*, in U. Ruffolo (ed), *L'intelligenza artificiale*, 72, «such systems do not provide explanations for their decisions», entailing inevitable risks for the subjects subjected to such decisions. The decision-making path followed by the machine is therefore obscure in its steps and cannot be predetermined, and this is even more evident in the case of the so-called non-deterministic

of decision-making in institutional contexts such as the court of law. In this case, relying on the 'opaque' reasoning of the algorithm in fact entails the risk of submitting to decisions that are unclear in their foundations or even vitiated by hidden biases. The latter in turn may derive from biases that lie upstream of the algorithmic decision-making process, i.e. in the data set on which the machine is trained. On a deeper level, accepting decisions based on automated and unintelligible decision paths also means renouncing the guarantee dimension inherent in the idea of the judgement as a reasoned decision³⁷⁰. If the deliberative process is not explainable, motivation as we understand it is no longer possible and the recipient of the decision will find himself deprived of the necessary instruments of control, relegated to a space of passive subalternity³⁷¹. Hence the growing importance given to promoting the explainability of artificial intelligence systems artificial intelligence systems, in order to make the processes underlying data processing and the resulting decisions accessible and comprehensible to man. In this sense, the solution would therefore be to bring the explanatory capacity of artificial intelligence systems closer to that which characterises human decision-making, transposing the possibility of explaining the determining factors and logical steps of the decision-making process into an algorithmic key. The intelligibility of decision-making mechanisms constitutes a decisive factor on the path towards a reliable artificial intelligence, which as such can work alongside man and assist him in his various activities. If this is true for the generality of the cases in which algorithmic decision-makers can be used today, it is even more so in the context of the legal system, more so in the jurisdictional context, where the decision - or rather the reasoned decision – constitutes a fundamental instrument of guarantee, which is recognised in Article 111 paragraph 6 of the Constitution³⁷².

In this regard, numerous critical voices have emphasised how the opacity of the algorithms is structurally incompatible with the duty to state reasons and is therefore at odds with our

algorithms, which starting from the same input can lead to different results». On this point, see M. PALMIRANI, *Big Data e conoscenza*, in *Rivista di filosofia del diritto*, 1, 2020, 73-92.

³⁷⁰ As emphasised by C. CASONATO, *Costituzione e intelligenza artificiale*, in *BioLaw Journal*, No. 2s, 2019, 721; the opacity of the algorithms does not allow one to reconstruct the logical-argumentative steps and thus the reasons behind the decision. Faced with this, decision-making activity loses its justificatory dimension, justification being understood as "a discourse of an argumentative nature designed to show, by means of reasons, that something in the field of doing [...] is right in a broad sense, that is to be accepted, preferred, chosen, pursued, precisely on the basis of reasons" according to the definition of U. SCARPELLI, *Gli orizzonti della giustificazione*, in L. Gianformaggio - E. Lecaldano (eds), *Etica e Diritto. Le vie della giustificazione razionale*, Roma-Bari, 1986, 12-13.

³⁷¹ The unknowability of algorithms thus represents an obstacle to the full expression of human control over machines and carries the risk of an unacceptable compression of the human rights involved in decision-making. Recalling the comparison evoked by F. PASQUALE, *The black box society*, 190; men in the presence of the 'black box' of technology are like the prisoners of the cave in the Platonic myth, destined to see only blurred shadows and therefore vulnerable to manipulation.

³⁷² On this topic also, C. CASONATO, *Costituzione e intelligenza artificiale*, 721.

system of jurisdictional guarantees. These perplexities, undoubtedly justified, can only fuel distrust towards the algorithmic judge, who, although he has on his side the efficiency and speed processing speed of the computer, seems destined to capitulate before the continuing protagonism of the human judge³⁷³.

11 When risk assessment approaches Artificial Intelligence

On closer inspection, the reason why we began the development of the paper by focusing on the characteristics of A.I. and then moved on to the study and analysis of risk assessment is because the latter have come to take on certain characteristics in this recent branch.

In fact, after having highlighted the main characteristics of risk assessments and their historical evolution, it is necessary to focus on the phenomenon of their now frequent and recurring translation into software; in fact, the same, although based on algorithmic models, also used in risk assessment, in some cases employ artificial intelligence solutions.

This now not-so-recent trend obviously adds a further level of complexity to risk assessment tools, which from the outset are structured in algorithmic terms as they superimpose, in the judgement of accuracy, the issue of digital opacity on that of the scientific validity of the psycho-criminological theory that inspires each tool³⁷⁴.

11.1 Two characteristics compared: accuracy and predictive significance

On closer inspection, risk assessment in the context we are dealing with, in particular in the psycho-criminological and clinical forensic context has short, medium and long-term implications as it has a considerable and direct impact on the quality of life of individuals and their families (here we speak of the microsocial aspect), but at the same time, it also affects the way the social, legal and scientific community responds to issues such as crime and violence and the practices put in place to manage, control and reduce them (also referred to as the macrosocial aspect).

In order to comply with the attempt to delineate a concept of predictive accuracy (valid for the assessments proposed here) it is considered useful and very necessary to understand what tools the scientific and clinical community possesses, in addition to those that the legislature

³⁷³ As emerges from E. FRONZA - C. CARUSO, *Ti faresti giudicare da un algoritmo? Intervista ad Antoine Garapon*, in *Questione Giustizia*, 4, 2018, 198, In the face of the advent of algorithms, the traditional conception of process and motivation comes into crisis, "because we prefer, even with all its weaknesses, a human authority over the automatism of machines".

³⁷⁴ On this point, S. QUATTROCOLO, *Artificial Intelligence, Computational Modelling and Criminal Proceeding*, 152.

makes available to the criminal justice system to directly affect the reduction of the risk of violence and crime and to intervene in an attempt to re-educate, rehabilitate and treat persistent and violent criminal individuals.

Undoubtedly, there are no perfect, golden instruments, the so-called golden tests, whose results are infallible and free from error; indeed, the outcomes of assessments must always be seen and understood as an indication of probability. Consequently, the resulting actions imply choices as to which risk one is most willing to take.

In the present case, the question arises and needs to be assessed as to what is preferable between assessing not at risk a person (e.g. by taking the decision and direction towards the benefit of early release) who will then commit a new act of violence; or, on the contrary, assessing at risk a person who will instead not commit any new violent manifestation.

This type of analysis refers to the concept of 'number needed to be treated' (NNT) analysis, which implies the estimation of the number of criminal individuals to be kept in detention and subjected to treatment in order to obtain a unit of advantage over those in detention and not treated.

It should also be noted that any risk assessment tool will always provide a certain proportion of false-positive and false-negative results. The dilemma behind these calculations is always that of choosing between a so-called conservative error or excess of false positives or type I error, or, on the contrary, a conservative error or tolerance of false negatives or type II error.

In order to provide and carry out an accurate risk assessment, indeed, it is not necessary to consider relevant those characteristics and events merely because they are present in the reality of the person under observation, but it is essential to be able to discriminate, with formalised and specific criteria of inclusion and exclusion, those that have only a descriptive value, from those that are instead criminogenically relevant for the purposes of violent or criminal relapse.

In this sense, it would be opportune and useful to refer to the importance of being able and knowing how to separate the signal (meaning all the information to be taken into account in any assessment) from the noise (meaning the biases that distract and influence our attention by polluting it with inaccurate elements that alternate the quality of the assessment).

In conclusion, it can be noted how the indices of sensitivity and specificity become the playground of any accurate assessment, where the former (sensitivity) implies the frequency with which true positives are reported; while the latter (specificity) indicates the frequency with which true negatives are recognised. Therefore, an evaluation is said to be sensitive if it has a low frequency of false negatives; conversely, an evaluation is specific if it has a low probability of false positives.

Consequently, an instrument that is very specific rarely misclassifies high-risk individuals, just as an instrument that is very sensitive does not misclassify non-risk individuals. Indeed, an optimal classification model is such when it is capable of maximising both sensitivity and specificity at the same time, but this is not possible or feasible³⁷⁵; it follows that such precise evaluations do not exist and that therefore any assessment of risk implies a margin of error and therefore a presence of false positives and false negatives. This can be explained by the fact that there is a trade off between sensitivity and specificity indices, which then leads to a chain of counter-reactions.

Instead, predictive significance as an index is closely connected in the use of risk assessment tools to being able to establish the risk level of those identified as problematic and their clinical usefulness in managing risk, reducing it, depends on a careful identification of the relevant risk factors and criminogenic needs, their adherence to scientific standards, compliance with procedures and ethical principles, and the methodology followed for the assessment. Indeed, even in risk assessment the assessment tools used must ensure that the predictive value of the assessment is able to correctly identify those persistent and recidivist offenders who are at high risk of criminal and violent relapse (the so-called positive predictive value VPP), but at the same time must also favour a correct discrimination of those who are not at risk of persistence and recidivism (even if they are considered criminals, the so-called negative predictive value VPN).

It should immediately be noted that the positive predictive value does not only depend on the test and its accuracy, but also on the prevalence of the risk of violence in the examined sample. Indeed, if the prevalence of violence is examined in a high-risk sample, this will influence the predictive value.

Finally, the reliability of these instruments must also be questioned: indeed, an instrument is reliable if it is able to offer the same result during repeated measurements when conditions remain unchanged. Consequently, the administration of the same test, given the repeated evaluations, should return the same results over several measurements and thus offer a reliable assessment.

An instrument can also be considered valid in its ability to distinguish high-risk and non-risk individuals in a population. It follows that an optimal instrument is one that is capable of being both very sensitive and very specific.

³⁷⁵ On this point, G. ZARA, *Tra il probabile e il certo*, 88.

Sensitivity remains the ability to correctly identify individuals at high risk; specificity, on the other hand, is the ability to correctly identify individuals not at risk. Within the same instrument, sensitivity and specificity are interdependent characteristics.

In conclusion, all this said and analysed, an absolutely precise distinction will in reality probably be impossible, since there are grey areas in which individuals who, despite exhibiting criminogenic features, do not reoffend, just as there are individuals who, despite the absence of criminogenic features, engage in a range of criminal and violent behaviour. Since not all recidivist criminal individuals will test positive, the resulting uncertainty creates an area of overlap in the results applied to all criminal individuals, recidivist or not.

11.1.1 The risks related to implicit bias

On closer inspection, an undoubtedly important aspect concerns the possible margins of error or falsifiability of the results issued by the predictive software (whose usefulness is closely connected to the empirical verifiability of the results).

Since it is a support tool for the judge, it would be left to the latter to assess whether or not the programme is able to offer outputs that are not tainted by calculation errors from the same cognitive biases that in some cases distort the judge's assessment.

Generally speaking, however, it can be said that error in these systems can arise from two macro-orders of reasons. A first error, which is that deriving from the human being, concerns the person who enters the information into the software database; indeed, the A.I. tool is still designed by man³⁷⁶. On closer inspection, as the structure of the algorithm itself traces, at its basis there is a selection of data to be entered into it as input, so that it cannot be excluded that, at this stage, the programmer makes mistakes or omits to enter certain relevant variables. This fallacy is also referred to as omitted variable bias and could in fact result in a miscalculation of the programme.

Indeed, the element of the programmer's discretion is central, and recurs throughout the entire design process, starting with the collection of data in the brainstorming phase, the selection of the relevant elements to be included in the dataset, and ending with the analysis of the results obtained.

In fact, the programmer is confronted with a series of tasks: he must first select the elements to be entered as input and consequently assign them a numerical value and establish the so-

³⁷⁶ Precisely in this sense, 'data and data-sets are not objective information but are creations of human design'. On this point, L. MALDONATO, *Algoritmi predittivi e discrezionalità del giudice: una nuova sfida per la giustizia penale* in *Riv. Trim. – Dir. Pen. Cont.*, 2, 2019, 401 ss.

called inferential rules; the latter are necessary and relevant since they are aimed at extracting (data mining), comparing (data matching) or profiling (data profiling) the data³⁷⁷.

Unfortunately, the risks of possible algorithmic discrimination reveal the existence of an inherent margin of error that is difficult to resolve. In this sense, the use of expert systems may only reveal an appearance of greater objectivity and impartiality. The only advantage, however, with respect to human prediction error is the fact that, at least the machine's error is easier to detect than that of man, since it is not inherent (unlike cognitive bias) in the psyche; in this regard, one notes how algorithmic distortions are easily recognisable, since they come to light as early as the testing phase, i.e. the evaluation of results by the computer programmer.

The possible risks inherent in the use of mechanisms configured as risk assessment, which have already been tried and tested in the United States, have already been highlighted, albeit only at an embryonic stage; moreover, the so-called selection of risk factors on which the algorithms are based and work has been examined. It was found, indeed, that they are not entirely impartial, but are based on stereotypes that are discriminatory. In fact, for this reason, the risk of so-called 'implicit bias' was highlighted: that is, on the one hand, where the input is found not to be completely neutral, the 'query' out would risk being influenced by a prejudice that could therefore lead to the discrimination of individuals or social groups; on the other hand, the algorithm, which is conceived and interpreted by a human being, may trivially reproduce unjustified social preconceptions³⁷⁸.

It is precisely on this point that an additional consideration appears necessary: indeed, the algorithm could 'reinforce' the so-called "implicit stereotypes" that are physiologically present in the person who has to make a judgement, thus increasing the risk of a criminal law approach of the type of perpetrator and enemy; indeed, if they turn out to be 'sophisticated algorithms' that can anticipate the behaviour of certain subjects, the risk would be to fall into a sort of Lombrosian theory - so dear to the Positive School - clearly in violation of the principle of equality³⁷⁹, offensiveness proper to the criminal law of the fact³⁸⁰, based on the fundamental

³⁷⁷ On this point, C. PARODI-V. SELLAROLI, *Sistema penale e intelligenza artificiale*, 47 ss.

³⁷⁸ S. QUATTROCCOLO, *Intelligenza artificiale e giustizia*, 6.

³⁷⁹ V. MANES, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, 12 – 14; C. CASONATO, *Intelligenza artificiale e diritto costituzionale: prime considerazioni*, in *Dir. pubbl. comparato ed europeo*, May 2019, 101 ss.

³⁸⁰ G. ROMANO, *Diritto, robotica e teoria dei giochi: riflessioni su una sinergia*, in *Diritto e intelligenza artificiale*, in G. Alpa, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, Pisa, 2020, 112; V. MANES, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, 12 – 14.

canon of art. 25, paragraph 2 Const. and of the principle of guilt, correctly understood as guilt for the fact³⁸¹.

12 *First Concluding Remarks on the Use of Risk Assessment in Criminal Justice: Towards Fair Treatment Justice?*

The descriptive analysis carried out in this chapter gives us an insight, if not an exhaustive one, into what is meant by risk assessment and allows us to see how, in reality, it is a large container that includes, within it, a varied number of tools that can be used in different fields.

The descriptive analysis carried out in this chapter gives us an insight, if not an exhaustive one, into what is meant by risk assessment and allows us to see how, in reality, it is a large container that includes, within it, a varied number of tools that can be used in different fields.

On closer inspection, although the new challenges of modernity lean towards new technological horizons, scientific knowledge falls into a realm of epistemological indeterminacy when it encounters the principles and rules of the criminal trial and criminal justice system, with the inevitable fallout of the dialectical criterion on the scientific method³⁸².

The main question on the *sidelines* of this new interaction between artificial intelligence and criminal law does not concern (it would otherwise remain a reductive question) the real danger of a substitution of the machine for the human, but rather one should perhaps reason about the quality and choice of data and the direction to be given in proposing application uses. It remains in the background to assess, as will be done in the remainder of the paper, the increasingly powerful and disruptive effects that these technological tools can have in the criminal trial that may entail a real structural paradigm shift³⁸³ in the rules on due process³⁸⁴.

³⁸¹ For instance, in the North American legal system, the use of algorithms for granting bail has created new channels of discrimination and fostered mass incarceration phenomena for the socially weaker and poorer groups. On this point, see M. ALEXANDER, *The Newest Jim Crow*, published in The New York Times, 8th November 2018. Instead of favouring 'a new normative dimension based on calculation, neutrality, and not on the subjectivity or free conviction of the judge. Justice would thus be better ensured by algorithms (neutral), rather than by human beings (subject to perceptions and subjective and unpredictable variables)'; see E. FRONZA, "Code is law", note to the book of A. GARAPON E J. LASSÈGUE, *Justice digitale. Révolution graphique et rupture anthropologique*, Paris, 2018; BURCHARD, *L'intelligenza artificiale come fine del diritto penale? Sulla trasformazione algoritmica della società*, 1932 ss.; S. G. MAYSON, *Bias In, Bias Out*, in *Yale law journal*, Vol. 128, No. 8, 2019, 2226 ss.

³⁸² See, F. I. GAROFOLI, *Il rischio inquisitorio negli strumenti di Intelligenza Artificiale*, in A. F. Uricchio-G. Riccio - U. Ruffolo (eds), *Intelligenza Artificiale tra etica e diritti. Prime riflessioni a seguito del libro bianco dell'Unione europea*, 455 s.

³⁸³ "Scientific revolutions are characterised by continuous and violent upheavals that suddenly overturn diverse paradigms and axioms", thus C. INTRIERI, *Neuroscienze e diritto: una Nuova teoria giuridica sulla mente, in Sistemi intelligenti/ a. XXII, No. 2, August 2010, 255* who reflects on the thoughts of T. KHUN, *La Struttura delle rivoluzioni scientifica*, 1962.

³⁸⁴ *Ibidem*, 459.

With regard to the field of justice, which is of interest in the present case, it was a question of identifying those tools that are most widely used at present, turning our attention to North American states.

In conclusion, in addition to these premises that will serve as an incipit for the propositional considerations that will be made in the next two chapters, it is considered appropriate to assess the initial implications also from a reading of foreign literature.

On closer inspection, the international literature shows that ERA instruments used in the prison and forensic psychiatric field demonstrate significantly higher reliability than simple clinical judgement in predicting recidivism of violent behaviour. Their field of use is progressively widening and includes support for judgments of degree of risk variability, general indications on the possibility of recidivism of offences that are not exclusively violent, and the planning of therapeutic-rehabilitation interventions 'designed' on the individual characteristics of the patient. These developments in turn pose a number of issues related to the unspecificity of the information obtained from the individual instruments with respect to what kind of risk of reoffending and to the 'middle field' effect, described by some authors, on subjects with intermediate scores but 'dangerousness' not different from baseline scoring obtained from empirical assessments. Actuarial scales, due to their static and unchangeable nature, appear to be of more uncertain applicability in the gradient from legal and prison settings to general psychiatric settings. The development of the most recent generation of scales based on a structured clinical approach accentuates the aspect of prevention of adverse events, heterodirected aggression and acted violence in the first place, but also other criminally relevant recidivism, downgrading to a secondary task the predictive and forecasting aspect, which are ill matched with the professional identity and competences of the psychiatrist and which raises not a few ethical questions. Risk analysis, if conducted as a tool of analysis involving the clinical unit, the social services, the forensic psychiatric area and the judicial system, becomes first and foremost a practice of systematic and qualitative collection of information from which the services involved draw common language, clinical investigation tools, ways of structuring management intervention, follow-up and audit. This necessitates, at the level of regional health policies and within individual departments, organisational adjustments that facilitate the fluidity of communication between the various operational units and the possible creation of transversal teams that are able to act as hinges between the various actors involved in individual cases. Adequate training should be provided on the notion of risk factors, the main ERA scales and their use, the dimensions of antisociality, psychopathy and their quantification. The training becomes a fundamental moment for acquiring awareness of

the area of investigation and application methodologies. In a crucial moment for the evolution of Italian psychiatry such as the one we are going through following the aforementioned Law 9/12, the introduction of knowledge and tools capable of bringing about quality improvements in the performance of the tasks to which professionals are already called upon, appears no longer to be derogable. We believe that the informed and conscious use of increasingly valid violence risk assessment tools can contribute to build an articulation of knowledge and interventions up to the complexity of the work dedicated to the psychiatric population in general and to the judicial population in particular: the ideal goal of a practice innervated by the knowledge we are writing about will be to facilitate the permanence in civil society, or the return to it, of subjects free from legal constraints and criminogenic needs.

The basic idea is therefore to move towards a justice that looks at the aftermath, at the post-sentence and at the way in which the Italian justice system can be readjusted in such a way that it becomes a real treatment justice, thus keeping the judicial level and the person's recovery level in a fair and balanced equilibrium.

Therefore, effective treatment practices require an assessment of both the risk and the criminogenic needs of the offender. In the event that such a diagnostic assessment is absent and no classification is made, criminal individuals enter the so-called 'treatment lottery'³⁸⁵, in which access to effective, targeted and specific, but above all personalised programmes is determined only by 'chance' and not influenced by internal and external correspondence.

This very rationale underpins the rationale and functioning of risk assessment tools.

In closing this chapter, we anticipate, as of now, what will be analysed in the following pages.

After having described the functioning and introduced the description of these tools, it was decided, from a scientific point of view, to assess how they can be applied according to and following a twofold direction.

On the one hand, and it will be dealt with in chapter three, an attempt will be made, starting precisely from the concept of social dangerousness, to understand precisely how these instruments could prove useful to the expert who has to make prognostic assessments that look to the future. In a second moment, and this will be dealt with in chapter four, an attempt will be made to understand precisely how these tools could prove fundamental to the adjudicating body in establishing and identifying the best sanctioning treatment.

³⁸⁵ G. ZARA, *Tra il probabile e il certo*, 81.

In order to obviate such drawbacks, one could hypothesise the preparation of 'programming protocols', which would be suitable for guiding the developer especially in the most delicate phase, i.e. that of assigning a numerical value to the individual elements of the data set.

The objective in signing such protocols could be to provide objective criteria and guidelines regarding the programming of the expert system, in order to reduce the risk of evaluation errors and to ensure compliance with the principle of non-discrimination.

Now, another type of error (somewhat more insidious and in some ways ineradicable) is that connected with machine learning systems³⁸⁶. Having assessed the risks of possible algorithmic discrimination, it has been pointed out how the reliability of the output may be affected by possible discriminatory effects due to the empirical generalisations and social and economic conditioning factors processed by the algorithm. Indeed, algorithmic distortion would lead to systematic and discriminatory errors, the so-called "algorithmic biases", causing a clear violation of the principle of equality, pursuant to Article 3 of the Constitution³⁸⁷.

³⁸⁶ This type of distortion has been described as follows by US mathematician C. O'Neil, who stated that 'far from being objective and transparent mathematical models, the algorithms that now dominate our hyper-connected everyday life are often veritable weapons of mathematical destruction: they do not take fundamental variables into account, they incorporate biases, and if they are wrong, they offer no possibility of appeal'. Thus we refer to C. O'NEIL, *Weapons of Math Destruction. How Big Data Increases Inequality and threatens Democracy*, Penguin Books Ltd. On this point also P. TILLERS-E.D. GREEN (eds), *L'inferenza probabilistica nel diritto delle prove. Usi e limiti del bayesianesimo*, Milan, 2003.

³⁸⁷ See, V. MANES, *L'oracolo algoritmico e la giustizia penale*, 547.

Chapter Three

A first perspective: the prognostic analysis of 'dangerousness'

SUMMARY: 1. Foreword: the concept of social dangerousness. – 1.1. The debate between the classical school and the positive school. – 1.1.1. A closer look at the Positive School: the delinquent's dangerousness. – 2. The structure of social dangerousness in the Italian Criminal Code. – 2.1. The passage and journey of the birth of the prognosis. – 2.2. Current prognostic criteria in different cases and their limitations. – 3. The two different types of danger. – 3.1. When the legal system finds itself making prognostic assessments. – 3.2. A decision looking to the future: between limits and difficulties of prognostic assessment. – 4. The judge's decision and the beyond reasonable doubt criterion in prognostic evaluations. – 5. Prognostic assessments referred to the judge in the Italian legal system. – 5.1. The different types of *prognosis* at the trial stages. – 5.2. Prognostic evaluation in security measures. – 5.3. Prognostic evaluation in prevention measures. – 6. The structural characteristics of the prognosis of dangerousness. – 7. How Artificial Intelligence intervenes in the judgement of dangerousness. – 8. The problem of defusing cognitive bias: possible advantages in the use of predictive algorithms. – 8.1. The second step in risk assessment tools: the individual and the group. – 8.2. The advantages of a mixed algorithmic evaluation: the US example. – 9. Possible remedies: enhanced and explanatory justification of the new algorithmic indices.

1 Foreword: the concept of social dangerousness

It should be noted that disciplines such as psychiatry, criminal anthropology, and forensic psychiatry have long and always strived to provide institutions with a certain answer regarding the dynamics that drive a person to commit a crime (capacity to understand and act) and regarding the dynamics that can drive a person to re-offend (determination of social dangerousness)³⁸⁸.

In the psychological tradition, there have been numerous attempts to link observed behaviour to individual characteristics, especially since such links could make it possible to explain and predict conduct with negative effects on individual and collective well-being. On closer inspection, the concept of dangerousness, a pivotal institute in the thinking of the Positive School, greatly conditions the entire original structure of the Rocco Code³⁸⁹; this is

³⁸⁸ Indeed, 'the relationship between psychiatry and law, in particular between psychiatry and criminal law, arose in the second half of the 19th century, when the psychiatric model was seen not only as a decisive paradigm for explaining otherwise absurd and inexplicable crimes in rational terms, but also as a possible model on which to base the new architecture of social control', T. SANNINI, *La genesi storia del concetto di pericolosità sociale*, 2014.

³⁸⁹ "The Rocco code was the first Italian regulatory source to entrust to the competence of the criminal judge the treatment of those subjects who, having been acquitted, were deemed socially dangerous. In the elaboration of the regulatory provisions functional to the new requirements, the Rocco Code was strongly influenced by the reformulation of the category of prevention and that of retribution elaborated by the Swiss-Germanic school, entrusted above all to Stoos and Exner, which, first forcibly separated general prevention and special prevention into two conceptually distinct aspects, and then confused the general aspect of prevention itself with the concept of retribution, thus transforming it into an instrument of crime prevention and depriving it of its function as a

explained by the fact that the notion of dangerousness, which - using the definitions given by commentators of the time - can be defined as "the power, aptitude, and capacity of a person to be the cause of harmful and dangerous actions and thus of damage and danger; more specifically, as the anthropological power apprehended in the offence, as the probable commission of further offences"³⁹⁰, has since then been placed at the centre of the debate on the 'control of the delinquent' (and at the centre of the debate on the double-track system).

Indeed, it is necessary to anticipate from the outset that social dangerousness - understood as the probability (and not mere likelihood according to the logic of the Rocco Code) that a subject who has committed a crime will, in the future, commit other forms of conduct envisaged by the law as crimes - has always (since the 18th century) been the focus of studies and research by jurists and scientists who have dealt with this matter, in order to attempt to isolate the different or individual variables capable of identifying the cause that led a subject to commit a crime. Indeed, the intention behind the study was precisely to elaborate a robust theoretical apparatus capable of providing the appropriate assessments - from a psychological and criminological point of view - in order to make a prediction that the subject will not commit further crimes.

Over time then (and the second point will be the focus of the analysis) the concept of 'social dangerousness' - provided for and regulated in Article 201, paragraph 1 of the Criminal Code, has become closely linked to that of the capacity to commit crimes, pursuant to Article 133, paragraph 2 of the Criminal Code, a difference and parallelism that has always determined an area of encounter between the legal and scientific fields³⁹¹.

It is, however, a fact that the category of dangerousness is reflected in the analysis of common experience by the presence of individuals who can potentially commit crime or return to crime; for the scholar, therefore, the distinct aspects of ascertaining and treating a subject (even a non-chargeable subject) who is concretely deemed socially dangerous emerge, thus expunging any form of legal presumption.

criterion limiting liability, typical of the Enlightenment's guarantee system', thus, again T. SANNINI, *La genesi storia del concetto di pericolosità sociale*, 2014.

³⁹⁰ E. ALTAVILLA, *Studi sul progetto del nuovo codice penale. Visione positivista della parte speciale di un nuovo codice criminale*, in "Scuola Positiva", 1921, 428. For more on the subject of dangerousness, see, N. PALOPOLI, *Il Progetto Ferri fra la Scuola positiva ed il moderno indirizzo criminale*, in *Scuola Positiva*, 1925, 13. In prospettiva critica, see A. CORDOVA, *Le riforme della legislazione penale e il loro momento storico*, in *Rivista Penale*, 1921, 101.

³⁹¹ On this point, see Chapter IV.

The central problem regarding this issue has always been that of the need to typify cases of dangerousness that are respectful of the principle of strict legality, provided for in Article 25, paragraphs 2 and 3 of the Constitution.

It cannot but be noted how, in reality, in certain penal sectors, the boundaries of the applicative marked area of certain instruments (such as, for example, the measures of prevention) have been far exceeded; in other cases, instead, the legislator has sought to outline more defined boundaries (one thinks of the case of the measures of prevention marked by the juvenile criminal law, in which an attempt has been made to individualize, as far as possible, the visions of judges and experts around solid points of reference of a criminological nature).

However, today we are witnessing, albeit against all attempts to delegitimise the phenomenon, an expansion of the category of dangerousness in the criminal justice system in order to prevent the risk of a proliferation of crime in the context of a society such as we have today.

Indeed, the criteria of 'positivistic memory' re-emerge as certain data of personal knowledge of a perpetrator can be predictive of future criminally relevant conduct. Undoubtedly, an exasperated search for the scientific nature of concepts (regardless of Popper's considerations on refutability and fallibility as intrinsic characteristics of human knowledge) partially veils the gaze of the criminalist who cannot blindly defer to the intuitive capacities of a judge who finds himself less and less ready to grasp the essence and complex meaning of dogmatic institutions and categories³⁹².

In order to explain the reasons why we question the concept of social dangerousness, it would first be appropriate to take a step back and ask why we question this concept and first of all, perhaps, to answer a question that has always interested the study of criminal justice. In fact, closely linked to the theme of dangerousness, which serves as an outpost to why we should question the second paragraph of Article 133 of the criminal code in particular, it would be necessary to ask 'why punish', why one should react with evil to one who has done evil and this is achieved through the instrument of the reinstatement of the legal order and therefore to ask within what limits this retribution should take place. However, we do not consider this to be the right place for such a reflection, which will, therefore, contribute to 'background' a delicate subject that therefore requires the utmost attention as it affects the punishment of an individual.

³⁹² "Hence, too, the legislator's commitment to shape, in the sector under consideration, cases that respect the canon of strict legality in the dimension of typicality and determinacy, so as to avoid any instrumentalisation in the use of the concept of social dangerousness", thus on the point, P. MAZZA, *Pericolosità sociale e legalità*, Milan, 2012, 3.

Such premises are considered to be more necessary than ever, since the judge today is faced with a very wide range of 'sanctioning responses' and different types of punishments to be imposed on a subject considered guilty; at the same time, he is faced with the need to issue both in the decisional phase (sentencing) and in the phase on the possible applicability of security and prevention measures, on the dangerousness of an individual or, more specifically, on the characteristics of his personality, which then lead to the decisional phase.

Therefore, starting precisely from a brief historical premise on the concept of dangerousness (as then delineated by the two principal Schools), an attempt will be made to understand how such an arduous and complex assessment, involves an evaluation that goes beyond the single fact committed and which, at the same time, addresses both retrospectively looking at the past and, in the choice on the measure or on the sanctioning treatment, at the future.

In conclusion, with these initial premises, it is essential to examine the doctrinaire positions on the subject of capacity to commit offences and dangerousness, with the need to go back, from the dispute on the subject, to the much deeper and opposing divergence existing between the two opposing schools, the positive and the classical, and to the philosophical components that constitute the roots of this divergence.

It appears useful to note how the two distinct figures of criminal capacity and dangerousness, since they pursue the same general objective, differ from each other as means that, by different routes and different immediate aims, tend to achieve it. The two figures therefore operate in a different specific field, even though they have peculiar connotations. As mentioned above, the assessment of the capacity to commit offences serves the purpose of commensurating the size of the penalty in the individual case and within the limits set by the legislation; at the same time, the assessment of dangerousness serves instead the purpose of applying, extending or terminating security measures. Indeed, social dangerousness operates fundamentally within the framework of correctional needs, having regard to the probable future conduct of the subject.

The evolution of the concept of social dangerousness and of the psychological and criminological theories mainly used by experts cannot be disregarded. Precisely for this reason, it is not possible today to rigidly define and circumscribe the concept of dangerousness, since it remains a relative principle and concept, closely linked to a specific historical moment experienced by a social context that raises particular alarm³⁹³. In part, it is therefore considered

³⁹³ On this see G. PONTI, *La abolizione delle presunzioni di pericolosità sociale*, in *Rivista italiana di medicina legale*, IX, 1987, 4.

difficult and almost impossible to give a scientific content and definition to the concept of social dangerousness, if understood in its original sense of prognosis³⁹⁴.

Lastly, it is necessary to assess, after the various oscillations between anchoring to the fact and to the perpetrator, since the 1930 Code, taking into account the instances expressed by modern criminalist currents and the evolutionary tendencies of criminal law, has extended its assessment from the fact to the perpetrator through, principally, the two new institutes of the capacity to commit offences and of social dangerousness.

In conclusion, in the following paragraphs, an attempt will be made to summarise how the theme of social dangerousness is pervaded by the evolution and alternation of the approach of the classical school and the positive school; furthermore, an attempt will be made to focus on the differences that have seen the two overseas and European positions contrasted and, above all, approached in different ways, and finally, to shift our attention to the evaluation and practical assessment of the particular form provided for in Article 133, paragraph 2 of the Criminal Code.

1.1 The debate between the classical school and the positive school

First of all, it must be premised that, with regard to the European territorial context, the phenomenon and the focus on the concept of social dangerousness reached the apex of the clash when it saw the opposition between the classical theory of punishment, based precisely on retribution, and, on the other hand, the positive ideology³⁹⁵.

The notion of 'social dangerousness' saw the light of day in the Italian legal system with the Rocco Code of 1930³⁹⁶. Indeed, as already mentioned in the introduction to this chapter, this notion has a vast and complex historical-ideological background, being at the heart of the long-standing and fiery controversy that, between the end of the 19th century and the first half of the 20th century, animated the debate between the Positive School and the Classical School of

³⁹⁴ However, in this scattered panorama that forms the backdrop to the discussion of dangerousness, there are also voices that define dangerousness as conceptually amorphous and in crisis insofar as it is based on inadequate and unclear predictive techniques and entails a particularly unfortunate mixture of 'therapeutic' and 'neutralisation' instances, which often finds concrete expression especially in the structure of the judicial psychiatric hospital, repository of the ambiguity connected with being both mentally ill (and therefore to be treated) and socially dangerous (and therefore to be neutralised). See, U. FORNARI, *Trattato di Psichiatria Forense*, Turin, 2004, 143.

³⁹⁵ In particular, with the German *Moderne Schule*, inspired by the concept of criminal law as a form of social control, which promoted a preventive purpose of punishment.

³⁹⁶ A code - the one signed by the then Lord Chancellor Alfredo Rocco - that does not present itself as a unitary whole but, on the contrary, succeeds in encompassing in a coherent system markedly authoritarian institutes together with principles that belong to the liberal nineteenth-century penal tradition. In this penal legislation, the influences of the Classical and Positivist schools have found an original synthesis, and with respect to these the Rocco Code has not failed to make its own original contribution. Thus, G.V. VASSALLI, *Codice penale*, in *Enciclopedia del diritto*, Vol. III, Milan, 1960.

criminal law. In fact, the problem of the free author, determined was tackled precisely in this period by these aforementioned strands and opposing trends in Italian criminal science, as well as by the mediating trend represented by the Third School³⁹⁷.

On closer inspection, particularly towards the end of the second half of the 19th century, a debate and controversy arose in the penal panorama that pitted the classical and the positive school against each other³⁹⁸. Indeed, and here we shall confine ourselves to hinting only at the essential points between the two contrasts that will be relevant to the analysis of the arguments discussed in this Chapter.

It can be seen from the outset, how the classical school³⁹⁹ placed at the basis of its Enlightenment-derived approaches the Aristotelian-theological principle of moral imputability, which was identified with free will and proclaimed and identified the law as the only possible source of incrimination⁴⁰⁰, invoking the canon of proportionality between the penalty and the

³⁹⁷ For a complete overview of the Schools mentioned, please refer to S. VINCIGUERRA, *Diritto penale italiano, Concetti, fonti, validità, interpretazione*, Vol. 1, Padua, 2009, 255; F. GROSSO, *Le grandi correnti del pensiero penale tra Ottocento e Novecento*, in L. Violante (ed) *Storia d'Italia*, Annali 12; AA. VV., *Scuola positiva e Codice Rocco*, in *Dir. Pen. XXI sec.*, 2011, 181.

³⁹⁸ See U. SPIRITO, *Storia del diritto penale italiano da Cesare Beccaria ai giorni nostri*, Turin, 1932, 23 ss.

³⁹⁹ The first attack on the classical conception of retributive punishment that saw culpability and the principle *nullum crimen sine culpa* as a guarantee principle, which presupposed freedom of the will, which was anchored to the individual criminal act and excluded that the judgement could extend to the offender's entire personality was brought by the criminological Positivism of the doctrine of purpose of V. Liszt, the most illustrious exponent of the Modern School, future founder of the International Criminal Law Union (1889), who, in the Marburg Programme of 1882, on the basis of a biopsychological determinist, and therefore ontological naturalistic view of crime, proposed to give a different legitimisation to punishment based on types of perpetrators and not on the fact. Remaining anchored to a monistic system that saw punishment as exhausting the range of criminal consequences, it declined the special-preventive function of punishment in its individualisation and differentiation. The penalty thus became a malleable and multifunctional means, superintending both neutralisation, intimidation and resocialisation, in relation to the concrete purpose that the individual case imposed. Although Liszt affirmed that the punishment thus conceived retained its retributive function, its practical outcome was essentially centred on social defence, hinging on a profoundly classist vision in the identification of irredeemable subjects to be perpetually neutralised (beggars, vagrants, prostitutes, alcoholics, scoundrels, degenerates in body and spirit) and on the offender before the crime. This theoretical approach could not but overwhelm the guarantee function proper to the classical conception, essentially manifesting correctionalist intentions of a paternalist and authoritarian type. (Musco, Ferrajoli) needs of criminal policy will cause this monist vision of a subjectivist type to find a compromise solution with the old conception of punishment (which will however lead to confusing the planes of external legitimation and internal legitimation of the same even on the strictly classical retributive level). For this reason, the idea will arise that in virtue of a functional limitation of punishment, a security measure with neutralising functions, a special preventive measure applicable to particular categories of offenders, will have to be added to it. Psychically abnormal subjects will be the classic paradigm to be taken into consideration.

⁴⁰⁰ Now, it should be noted that the main exponent of the Classical School is Francesco Ferrara, who wrote one of the most significant works of this orientation; according to the author, in fact, punishment is the imputation based on free will, which is assumed for the distinction between law and morality. If for any reason this freedom is lacking in the offender, criminal law is no longer applied to him, and if imputability is diminished, the penalty is also diminished. On the thinking of the members of the classical school, we refer to an interesting reading of G. BATTAGLINI, *Principi di diritto penale in rapporto alla nuova legislazione. Questioni preliminari*, Milan, 1929, 48 ss. Furthermore, according to R. DELL'ANDRO, *Il dibattito delle scuole penalistiche*, in *Arch. Pen.*, 1958, 173 ss., the Classical School extols rational data, while positivists advocate naturalistic-individual data.

offence⁴⁰¹; Moreover, it remained firmly in the conviction of this strand that offences should be described in delimited and precise cases, recognising that the punishment had an eminently retributive function, on the inalienable assumption of the offender's free will, with the consequent proportioning of the punishment itself to the gravity of the offence and guilt. On the contrary, the positive school, on the other hand, born in a totally different historical and social context, since it started from the consideration of crime as a human fact, a symptom of a given personality and characteristics, founded the penal system on the need to safeguard associated life, the only positive justification for the right to punish⁴⁰². He contrasted a completely different penal system that hinged on types of offenders and penalties, inspired not by a retributive criterion but by the idea of responding to a social defence. In particular, it was precisely this latter approach, embodied in the idea of human error, that completely rejected the principle of 'penal dosimetry'⁴⁰³, which was at the basis of the classical School, according to which the State is obliged to punish those who do wrong in order to maintain and preserve the legal order at a given historical moment, since society and law are two 'correlative and convertible' terms. Indeed, it is precisely with the Positive School that the idea that the penalty should not be commensurate with the seriousness of the offence or even with the prohibition and precept violated, but exclusively with the fearfulness of the offender, is strengthened and forged. In particular, it is a criterion that serves as a 'guide' for the application of the definitive suitable means, as a yardstick for the adaptation of the offender to the environment and such as to lead to the examination of the conditions of existence in which it may be presumed that he ceases to be fearful⁴⁰⁴; that means, in particular, is then to be sought and determined only as an outcome of the examination of the conditions of existence in which it may be presumed that the offender is no longer (again) fearful. Therefore, following this reasoning, the most suitable means should be sought so that the offender no longer represents an element of disruption⁴⁰⁵.

⁴⁰¹ On this point, authoritative is the position of F. CARRARA, *Programma del Corso di diritto criminale*, Lucca, 1867, §§ 659 and 697; Furthermore, for a peculiar approach that questions the fundamental principles of the classical school, see a G. BOVIO, *Saggio critico del diritto penale*, Naples, 1883, 37 ss.

⁴⁰² On this point and, in particular, on the inductive-experimental method of investigation, typical of the positivist orientation, and on the abandonment of the abstract procedures of syllogistic logic, which favour the study of crime as a biopsychic and social phenomenon, see E. ALTAVILLA, *La vitalità della scuola positiva*, in *Scuola pos.*, 1947, 77 ss.

⁴⁰³ On this point, E. FERRI, *Sociologia criminale*, vol. I, Turin, 1929, 15 ss.

⁴⁰⁴ On this point, see R. GAROFALO, *Criminologia. Studio sul delitto, sulle sue cause e sui mezzi di repressione*, Turin, 1861, 330.

⁴⁰⁵ "Everything therefore comes down to the determination of social necessity. General criteria and rules of application can only start from here. Everything that is attempted outside this chapter fatally leads to scientific errors which, translated into laws, turn to the detriment of society", thus on the point, R. GAROFALO, *Criminologia*, 330.

Given these brief remarks on the major differences between the Classical School and the Positive School, subsequent studies on the subject saw the gradual blurring of the objective and external profiles linked to the concept of fearfulness, to see instead a greater emphasis on the purely subjective aspects of assessing the delinquent's tendencies towards crime or his own re-socialisation⁴⁰⁶.

Following this approach, therefore, it is immediately apparent how the dangerousness of the offender stands as a true fundamental subjective paradigm that must replace the classic objective paradigm of the extent of the crime. It remains true, however, that dangerousness contains within itself, as its own intrinsic consequences, on the one hand, the greater or lesser fearfulness and, on the other, the greater or lesser adaptability of the subject to social life: while on the one hand the former retains a more direct scope in the security police, on the other hand, adaptability is more strictly adhered to the practical purposes of criminal justice⁴⁰⁷.

On closer inspection, faced with this composite and complex historical and cultural situation, the legislator arrived at a sort of compromise, in which, alongside traditional institutions of the classical conception, certain new and significant solutions adopted by the positivists were at the same time incorporated⁴⁰⁸.

In conclusion to these introductory and descriptive premises, we realise that for matters of method it is not possible here to dissect the subject and the debate not only legal but also philosophical that has marked the whole of the last century on this issue and this alternation and contrast of views.

However, what we would like to emphasise and clarify is that the confrontation between the two schools has in fact never 'stably' polarised into the binomials 'punishment/incapacitation/' and 'retribution/danger', since, in reality, although they are contrasting, they are not opposites but, inevitably, intertwined.

⁴⁰⁶ See, G. GUARNERI, *Pericolosità sociale*, in *Noviss. Dig. It.*, vol. XII, Turin, 1965, 952.

⁴⁰⁷ These reflections also concerned the contrasts that arose within the Positive School and the controversy, in particular, that matured within positivism between Turati and Ferri: the former, an advocate of a progressive 'decay' of criminal law; the latter, on the other hand, turned to the idea of removing the so-called 'social question' from the central role in the analysis of the various factors of crime and to evading the question as to why crimes are increasingly growing in capitalist societies, on the assumption that upstream of the phenomenon of crime, the phenomenon of crime is growing in capitalist societies, and that the 'social question' is not a central issue in the analysis of the various factors of criminality and to evade the question as to why crimes are growing more and more in capitalist societies, on the assumption that upstream of the criminal phenomenon there are various criminogenic elements and not only misery and economic exploitation.

⁴⁰⁸ Refer to G. DELL'OSSO, *Capacità a delinquere e pericolosità sociale*, Milan, 1985, 46 ss.

1.1.1 A closer look at the Positive School: the delinquent's dangerousness

It should be pointed out from the outset that it is extremely useful for the analysis and the continuation of the paper to focus attention on the category of dangerousness and the meaning that the Positive School has given to it. This passage is considered useful and necessary because positivist thought, which had, despite the liveliest polemics and debates triggered, a wide echo throughout the world, had a great influence on the evolution of criminal law and criminology⁴⁰⁹. Indeed, it may be useful because it concerns precisely the declination of the concept and the links with the choice and certainty of the penalty to be imposed on the convicted person. Indeed, the Positive School⁴¹⁰ (whose progenitor can be recognised in Lombroso and as other significant exponents Ferri and Garofalo) - albeit with different lines of approach - always demands that the criminal defence against crime be implemented by adapting the sanction to the offender's dangerousness on the basis of the canon of the so-called individuation of⁴¹¹. This approach has its philosophical and cultural roots in Methodological Positivism, which developed in the 19th century in opposition to Enlightenment rationalism.

The basic idea of the School⁴¹² consisted in the enunciation of the conception that punishment should necessarily be commensurate with the fearfulness of the offender; punishment did not merely pursue the aim of 'intimidating with the threat of an evil (in which the general preventive function of punishment is substantiated), which is expressed as 'direct general' prevention, deterring the evil-doers, and as 'indirect general prevention', which instead gives the honest the confidence in the efficacy of criminal justice, but also essentially the aim of 'preventing a new crime in one who was already capable of committing a crime in order to implement special prevention.

In addition, the conceptual framework constructed by the Positive School is therefore based on certain pillars: the focus is shifted from the crime to the offender; since free will has no

⁴⁰⁹ It should be recalled that the Positive School also has ideological-political matrices in a socialist or Marxist sense, as is evident from the fact that the innovations it proposed influenced both Soviet criminal science and Russian criminal legislation itself in 1922 and 1926. Thus, on this point we refer to G. BETTIOL, *Il problema penale*, 1945. However, the total alienation of penal socialism from positivist thought was subsequently demonstrated. In fact, the adherence of Soviet penal legislation to positivist postulates (especially with regard to the substitution of legal responsibility for culpability in an ethical sense and of social defence measures for punishment) reflects, albeit in the common 'basic utilitarianism', a convergence of more and different polemical positions against the individualistic and liberal instances of the classical school than of ideological orientations. Thus, on this point, refer to a L. PORZIO, *Sistemi punitivi e ideologie*, Naples, 1965, 20.

⁴¹⁰ The birth of the positive school is conventionally traced back to 1876, The position of Lombroso first, and then of the other representatives of the positive school, stems from the criticism of one of the pivotal concepts of the classical school, namely free will.

⁴¹¹ R. A. FROSALI, *Sistema penale italiano*, Vol. I, Turin, 1958, 35 ss.

⁴¹² In particular, the idea belongs to Raffaele Garofalo, one of the founders of the School, R. GAROFALO, *Criminologia*, 426 ss.

place, the concept of imputability is elided and the idea of social dangerousness becomes central, which includes both the possibility that a subject, due to his peculiar characteristics, commits another crime, and the intrinsic dangerousness of the person who has committed or attempted to commit a crime. The consequence of this conception is that security measures must be applied alongside the penalty or as a substitute for it.

Thus, following this starting point and elaboration, the criterion of the dangerousness of the offender came to represent the application parameter for punishing an offence, in complete and total opposition to the classical thought of the proportion between punishment and offence⁴¹³.

The central point on which the idea of the positivist School insisted and was based was the absolute prevalence of the consideration of the dangerousness of the offender, which was then delineated and demarcated following two directives: the social dangerousness ante delictum (which then opened up to a preventive assessment of dangerousness) and in criminal dangerousness, which instead gave rise to an appreciation of a repressive character post delictum. In this case, the importance of the crime in itself, which determines the intervention of the repressive penal instrument, was not denied, but at the same time emphasis was placed on the crime actually committed by the subject as a symptom revealing a socially dangerous personality. The positivists placed the dangerousness of the offender in the foreground, so as to enclose the unfolding of that justice within the four different phases of the crime, the offender, the judgement and the punishment⁴¹⁴.

The Positive School did not deny the importance of the crime in itself, but added to it its value as a symptom from which emerges a personality (to be assessed) that may be socially dangerous⁴¹⁵. The crime, considered before a legal fact, as a human and social fact, leads to the consideration of the evaluation of man and his relative relationships within society in all its interest; so that the individual is called to answer for his criminal actions not so much for a

⁴¹³ On this point, R. GAROFALO, *Criminologia*, 426 ss. In contrast to Garofolo's view, which took up an idea already expounded by Feuerbach, the same classical school had asserted that the danger was 'essentially circumscribed to the objective entity of the crime, considered as a legal entity referable to a man as a responsible subject and deserving of the punishment-castigation'. Any assessment of the person as such for the purpose of determining the sanction in concrete terms therefore remained at the margins.

⁴¹⁴ The different 'stages' of crime had been accurately marked out by one of the proponents of the Positivist School, Enrico Ferri, who had keenly noted how, while in an earlier period the crime was punished in the offender, in the future, the offender would be judged in the crime.

⁴¹⁵ And it was precisely from this element that the denial that the criminal act, if pondered and considered in isolation, could provide the full measure of the danger to be criminally appreciated derived.

moral judgement on his culpability⁴¹⁶, but for the unique circumstance of living in the organised community⁴¹⁷.

It is considered that the considerations made in this paragraph may serve as a premise in line with the direction taken in this Chapter; indeed, since starting from the broader question of trying to understand why dangerousness comes to the fore, it is essential to start from the concept of the same and to understand how the latter has changed over time. In particular, in an attempt to understand why the same remains, to this day, a fundamental concept because it embraces, from the point of view of the application consequences of the instrument of criminal law, both the phase of the assessment of dangerousness for the purposes of the recognition of security or prevention measures, but also, and not transversally, the analysis of the subject in the choice of the best sanctioning treatment. And it is precisely on this last point that the analysis in the following Chapter will focus.

Undoubtedly, one of the merits acknowledged to the Positive School is that of having focused on the problem of the delinquent's personality in its bio-psycho-sociological conditioning; furthermore, having understood the crime and the offender within the individual and social reality, then giving rise to the criminological, anthropological and sociological directions that contend with the field of criminology.

2 *The structure of social dangerousness in the Italian Criminal Code*

The current penal panorama in Italy undoubtedly offers a face of social dangerousness, as the probability of reiteration of offences, which expands beyond the confines of the codified tradition that sees social dangerousness as the main *prerequisite* for the application of security measures⁴¹⁸. And it is precisely in this expansion that dangerousness, already in itself a rather generic concept, since it is not anchored to the objective parameters of the seriousness and type of the predicate offence or those whose future commission is feared⁴¹⁹, increasingly becomes

⁴¹⁶ It is no coincidence that for the proponents of this school, the investigation of the motives that guide human conduct acquires decisive importance. On this point, C. LOMBROSO, *Trattato antropologico sperimentale dell'uomo delinquente studiato in rapporto alla antropologia, alla medicina legale e alle discipline carcerarie*, Milan, 1876.

⁴¹⁷ E. FERRI, *Lezioni*, Rome, 1911, 76 ss.

⁴¹⁸ Moreover, it is implicitly present as a justificatory rationale for institutions, such as recidivism, the ambiguity of which has long been highlighted by the doctrine, see, among others, M. BERTOLINO, *Il reo e la persona offesa. Il diritto penale minorile*, in *Trattato di diritto penale*, diretto da C. F. GROSSO – T. PADOVANI – A. PAGLIARO, *Trattato di diritto penale*, Milan, 2009, 137 ss.

⁴¹⁹ Not so for social dangerousness with reference to minors, cf. Article 35 of Law 689 of 24 November 1981. On the relationship between fact and dangerousness, see T. PADOVANI, *Fatto e pericolosità*, in *Pericolosità e giustizia penale*, in M. Pavarini – L. Stortoni (eds), *Pericolosità e giustizia penale*, Bononia, 2013, 117 ss.

a type of qualification of a strictly subjective nature, i.e. determined by the type of perpetrator and the social alarm that the same figure arouses⁴²⁰.

On closer inspection, the 1930 Criminal Code did not fully adhere to the direction of either school (although it was permeated more by the influence of the Positive School). Although permeated in the definition and framing of certain institutions by the logic of social defence, the code takes a middle position on the subject of social dangerousness. It follows, indeed, that criminal imputability (provided for in Article 85 of the criminal code) stands on the intermediate conception of psychic normality, identified with the doctrine of psychological determinism or psychic causality⁴²¹.

Indeed, the Rocco Code⁴²², although it remained anchored to the objective fact in the formulation of the incriminating case, undoubtedly denotes the extension of the assessment of the fact to the perpetrator through the two institutes of the capacity to commit offences⁴²³ and of social dangerousness. In fact, the typified figures of social dangerousness that are still present in our Code today are undoubtedly the result of the attention and interest shown by the positivist current to the subject understood as a delinquent (hence maximum attention to the author of the crime) as opposed to the attention paid to the crime and to the category of dangerousness; this choice responded to the more general need to create an ad hoc criminal sanction that takes into account not only the greater or lesser seriousness of the crime but also the greater or lesser dangerousness of the delinquent⁴²⁴.

⁴²⁰ In fact, as punctually observed, from the 'angle' of the 'individual aptitude for committing crimes ... dangerousness constitutes a category that, historically and ideologically foreign to the retrospective gaze of punitive retributionism, develops the securitarian aspiration of criminal law by renouncing not only the protection of guilt, but also that of offence', see F. GIUNTA, *Verso una nuova pericolosità sociale*, in *Cultura e diritti*, 2012, 3, 93.

⁴²¹ On this point, G. MARINI, *La capacità d'intendere e volere nel sistema penale italiano*, in *Riv. It. Dir. Proc. Pen.*, 1961, 733 ss. Furthermore, on the different theories concerning the basis of imputability, see G. CERQUETTI, *L'imputabilità nella sistematica del diritto penale*, Perugia, 1970; M. BERTOLINO, *Profili vecchi e nuovi della imputabilità e della sua crisi*, in *Riv. It. Dir. Proc. Pen.*, 1988, 202 ss.

⁴²² In the previous codified discipline, social dangerousness was not the subject of a prior ascertainment by the judge; on the contrary, it was the law itself that provided, in specific cases preemptorily identified, the hypotheses of presumption of dangerousness against a subject. These were presumptions *juris et de jure* which, as such, did not admit contrary proof. On this subject, we refer to some important decisions: Corte Cost., 20/01/1971 no. 1, in *Giur. Cost.* note by Vassalli, in which the court declared illegitimate the presumption of dangerousness of a minor who could not be charged; Corte Cost. 27/07/1982, no. 139, *Riv. It. Dir. Proc. Pen.*, 1982 p. 1585 with a note by E. Musco, in which the court declared illegitimate the presumption of dangerousness of the person acquitted on grounds of insanity (art. 222, par. 1, Criminal Code); Corte Cost. 28/07/1983, no. 249, in *Riv. It. Dir. Proc. Pen.*, 1984, 460, with a note by Giuri, in which the court declared illegitimate the presumption of dangerousness of the semi-normal person (art. 219 c.p.).

⁴²³ This topic will be dealt with in Chapter 4.

⁴²⁴ In fact, it was emphasised, it could well be the case that a serious crime was committed by a not very dangerous offender and that, on the other hand, a minor one was the symptom of an abnormal and very dangerous personality. In this perspective, repressive measures had to be more severe and effective for habitual offenders and less strict for occasional and, therefore, less dangerous offenders. Thus, on this point, see E. FERRI, *Relazione sul progetto preliminare di Codice penale italiano*, in *Scuola Positiva*, 1929, 5.

A peculiar aspect, already introduced by this code, is inherent in the part concerning the penalty framework: in fact, the discipline concerning the application and execution of penalties has been innovated; in particular, the judge is required to have a complete and organic vision of the offender and the crime committed by him, so that the penalty to be applied in concrete terms constitutes the result of a balanced assessment of the offender's personality and the seriousness of the crime⁴²⁵.

For this reason, the judge is required to take into account, when deciding on the determination of the punishment in concrete terms, not only the seriousness of the offence but also the offender's capacity to commit a crime (according to the indications provided for in Article 133 of the criminal code. At the same time, the types of criminological positivism of tendency criminals and habitual or professional offenders are also identified and regulated; moreover, certain fundamental institutions are better regulated than in the past in order to adapt them to the dangerousness of the individual and, above all, defence and re-education measures are provided for any person who has committed an offence provided for by law.

The concept and definition of social dangerousness make it possible from the outset to distinguish it from culpability. While on the one hand, the latter presupposes a sufficient 'sphere of lordship' in the individual to carry out his actions, social dangerousness, on the other hand, reflects the set of inclinations that drive the subject to commit crime in a necessary manner⁴²⁶. It should also be noted that in the previous codified discipline, social dangerousness was not the subject of a prior ascertainment by the judge, but it was the law itself that provided for the hypotheses of presumption of dangerousness of a subject. Indeed, these were precisely presumptions *iuris et de iure* which therefore did not admit of contrary proof⁴²⁷.

If we take into consideration the evolution of the concept of social dangerousness that has changed over time, of the psychological and criminological theories mainly used by experts, we can see how the concept of dangerousness⁴²⁸ and the qualification of dangerous offender

⁴²⁵ V. MANZINI, *Trattato di diritto penale italiano*, Vol. III., Turin, 1981, 249.

⁴²⁶ G. FIANDACA - E. MUSCO, *Diritto penale. Parte generale*, Bononia, 2011, 813.

⁴²⁷ The situation has profoundly changed thanks to Article 31 of Law No. 663/86 (the so-called Gozzini Law), which abolished all forms of presumption of social dangerousness in the code, repealing Article 204 of the Criminal Code. The law was the result of a series of pronouncements by the Constitutional Court, which declared illegitimate, over the years, all forms of presumption of social dangerousness, deeming them to be in clear conflict with the principles of the Constitution. On this point, M. CANEPA – S. MERLO, *Manuale di diritto penitenziario*, Milan, 1991.

⁴²⁸ For many practitioners, therefore, it remains impossible to give a scientific content to the question of social dangerousness, if understood in its original meaning of prognosis. There are those who do not hesitate to define dangerousness as conceptually amorphous and in crisis insofar as it is based on inadequate and unclear predictive techniques and entails a particularly unhappy mixture of 'therapeutic' and 'neutralisation' instances, which often finds concrete expression above all in the judicial psychiatric hospital structure, repository of the ambiguity

remain to all intents and purposes relative principles and 'qualifications', since they are accepted or rejected depending on the different degree of moral culpability or social alarm that a given event arouses in various historical moments⁴²⁹.

Today, in fact, for the application of a security measure, social dangerousness must always be the subject of concrete ascertainment by the judge. In fact, all the previous presumptive cases can only be considered as circumstantial hypotheses of a possible (but not certain) dangerousness⁴³⁰, thus fuelling the process of overcoming the infirmity-dangerousness binomial⁴³¹.

In conclusion, given, albeit in part, the evolution of the concept of dangerousness over time and the considerable influences that have determined the studies of the Positive School, the creation of the socially dangerous offender highlights the conception of a type of offender that is only partly abstract and partly concrete; in the abstract, since there is a type of dangerousness that is only presumed by the law, and in the concrete, since in cases of presumed dangerousness the judge is allowed, within certain limits, to assess it.

And it is precisely on this last point on which we want to move the line of this research, in an attempt to investigate in the first place the characteristics of the dual concept of dangerousness to crime in order to evaluate abstractly what are the characteristics that the same code takes into consideration and, in a second moment, having assessed the space for discretionary *manoeuvre* entrusted to the judging body, attempt to assess how the same can be usefully supported by an artificial intelligence tool.

2.1 *The passage and journey of the birth of the prognosis*

As can be seen from the evolution of the concept of dangerousness, while on the one hand the unitary model of indeterminate punishment was gradually expanding in the United States,

connected with being both mentally ill (and therefore to be treated) and socially dangerous (and therefore to be neutralised). On this point. U. FORNARI, *Trattato di Psichiatria Forense*, 143.

⁴²⁹ About this point, see G. PONTI, *La abolizione delle presunzioni di pericolosità sociale*, in *Rivista italiana di medicina legale*, IX, 1987.

⁴³⁰ G. PONTI - I. MERZAGORA BETSOS, *La abolizione delle presunzioni di pericolosità sociale*, in *Riv.it. Med. Leg.*, IX, 1989, 18 ss.

⁴³¹ Indeed, the evolution of social, psychiatric and legal sciences has made it possible to ascertain, at the very least, that 'the mentally ill do not commit crimes to a greater extent than the rest of the population'. It seems important to emphasise this concept especially in the light of Law No. 81 of 30 May 2014. It established that, in order to ascertain the social dangerousness of a mentally ill or semi-injured person, the requirement set out in Article 133 paragraph 2 no. 4 of the Criminal Code - i.e. the offender's individual, family and social living conditions - should no longer be taken into consideration, but only the person's subjective qualities should be taken into account. Moreover, the aforementioned law expressly provides that 'the mere lack of individual therapeutic programmes does not constitute a suitable element to support the judgement of social dangerousness'. Thus, on this point, see U. FORNARI, *Trattato di Psichiatria Forense*, 142 and also F. SCHIAFFO, *La pericolosità sociale tra "sottigliezze empiriche" e "spessori normativi: la riforma di cui alla legge n. 81/2014*, in *DPC*, 14 ss.

as already mentioned, on the other hand, it was precisely in that environment that studies on the analysis of prognostic judgments found favour, which were now placed at the centre because they were considered to represent a sort of 'keystone' of the entire repressive apparatus.

On the contrary, in Europe, the new horizon that had been opened up in the wake of criminological positivism found a compromise solution.

In fact, the world of punishment remained hooked on the idea of retributive proportion and consequently immune to any knowledge of the offender's future behaviour.

Special prevention and the prognosis of the latter have chosen the system of security measures as their privileged field of action. Indeed, the double sanctioning system devised by the Rocco Code, in part, perfectly reflects this compromise: in fact, in order to cope with the social dangerousness of the offender, whose prognostic assessment was based on legislative presumptions, the duration of the security measures was made indefinite in partial implementation of the programme formulated by criminological positivism. However, despite this, no action was taken to attempt to refine the prognostic instruments that until then had remained confined to the binary of social dangerousness and security measures.

2.2 *Current prognostic criteria in different cases and their limitations*

On closer inspection, there are provisions in the substantive code and in the code of criminal procedure that show the clear legislative attitude and intent to provide the judge with guidelines and guidelines when he finds himself making and issuing prognostic assessments.

Indeed, on the one hand there is the case provided for in Article 274 of the Code of Criminal Procedure and on the other hand the case provided for in Article 133(2)⁴³² of the Code of Criminal Procedure (both provisions closely linked to Article 187 of the Code of Criminal Procedure).

These cases, however, have a twofold function within them: on the one hand, they fix the themes of evidence for the parties, and, at the same time, they outline the motivational outlines for the judge, who is required to provide a basis for the assessment of an individual's dangerousness.

Despite the fact that the two provisions are to be found in two different regulatory sources and touch on institutions and decision-making aspects that are distant from each other, it can be seen that in reality some of the elements described by the standards are partially coincident: in fact, in the assessment of both cases, 'it is necessary to verify the modalities and

⁴³² On this point, Chapter 4.

circumstances of the fact, as well as the character or personality of the defendant, inferred "from his behaviour or concrete acts or from his criminal record" (according to art. 274, paragraph 1, lett. c, c.p.), or from criminal and judicial precedents and, in general, from the conduct and life of the offender, prior to the offence; from conduct contemporaneous with or subsequent to the offence; finally, from the individual, family and social conditions of life of the offender (Article 133, paragraph 2, c.p.)".

In particular, these data are mostly intended to be 'objective' but are considered to possess an ineliminable subjective component.

3 *The two different types of danger*

The 1930 Criminal Code had already taken care to identify two forms of dangerousness: one generic or simple and the other specific or qualified. In fact, the first type of dangerousness is reflected in Article 203 of the Criminal Code, where the first paragraph states that 'for the purposes of criminal law, a person is socially dangerous, even if not chargeable or not punishable, who has committed any of the acts indicated in the preceding article, "when it is probable that he will commit new acts envisaged by the law as offences". The next paragraph states that the quality of socially dangerous person is inferred from the circumstances indicated in the preceding Article 133 of the criminal code.

Indeed, this provision shows, on the one hand, the rejection of the positivistic postulate that wanted the judgement of dangerousness to be detached from the commission of a fact abstractly configured as a crime; on the other hand, that the essence of dangerousness can only consist in the probability that the subject will commit crimes in the future"⁴³³.

It appears more necessary than ever to take note that in recent years we are witnessing the crisis of this prohibition; indeed, the jurist today cannot fail to examine the more explicit and at the same time more underlying critical issues concerning prognostic judgements. Indeed, one cannot fail to realise that these necessary evaluations, which concern various choices entrusted to the judging body, probably 'ask too much' of a judge who cannot avoid basing or influencing his prognostic decision on his intuition.

Therefore, such evaluations would be tainted by a judgement that is not only personal, but increasingly disconnected and distant from the objective elements with which it can be compared.

⁴³³ B. PETROCELLI, *La pericolosità criminale e la sua posizione giuridica*, Padua, 179 ss.

On closer inspection, therefore, this prohibition is in crisis in the face of the irruption of the new technologies that can offer the jurist and also the judge new tools with which to compare and improve assessments.

On closer inspection, social dangerousness is a generic and constantly evolving concept that encompasses a multitude of meanings.

Today, the judgement of capacity to commit offences is necessarily projected into the future for the purpose of a prognosis of the probability of committing another offence or an offence of the same type, but it does not necessarily have predictive value.

Thus, the assessment of the offender's character requires a complex evaluation of his personality and of the innate characteristics of the subject that are capable of guiding his behaviour (e.g. his capacity for self-control and emotional stability).

The subject's social and family conditions (e.g. marginalisation, unemployment) are also subject to assessment. Moreover, as already noted or mentioned in the previous paragraphs, the subject's social and family conditions are also subject to assessment and, therefore, precisely because of the complexity of the factors to be assessed, it has been decided that the judge should carry out the examination independently, on the basis, however, of a rational and logical reasoning that takes into account the offender's frailty, without resorting to technical-scientific assessments.

Thus, in fact, the algorithm provides greater certainty and objectivity than human evaluations only apparently. In fact, the quality of the data entered and the statistical correlations would be 'flawed' by bias and would risk providing a result that does not adhere to reality⁴³⁴.

3.1 When the legal system finds itself making prognostic assessments

It is already anticipated here that there are at least two main areas in which most Western legal systems link a judicial decision to a 'prediction' of the defendant's future conduct.

One is undoubtedly that of the application of precautionary measures and the other that of the commensuration of punishment⁴³⁵. These two very delicate moments entrusted to the judicial body, reflect the basic principles, acting as a mirror of the guarantees' line of a penal system. This is because, intervening as a limitation or extension of personal liberty (Art. 13 of the Constitution), they are somewhat the mirror and the implication of a penal system marked

⁴³⁴ As noted «even with masses of data, there is no automatic technique for turning correlation into causation. See, D.J. SPIEGELHALTHER, *The Future lies in Uncertainty*, in *Science*, 2014, Vol. 435, 264.

⁴³⁵ Please refer to Chapter 4.

by different principles. As can be seen, in fact, arrest and pre-trial detention on the one hand and punishment on the other are closely linked to the criminological theories that have developed over time⁴³⁶.

As can be seen, both types of decision, although they intervene at different procedural moments, involve a risk assessment to be carried out on the individual looking to the future.

As already mentioned, in fact, it is an 'endo-procedural' risk, which has an impact on the conduct of the proceedings, or instead, it can have and assume a 'social risk' such as that of recidivism, following the execution of the sentence.

Indeed, the need to incorporate a risk assessment into judicial decisions on caution and sentencing could be the key to understanding how important the prediction of an individual's future behaviour can be in the context of a type of assessment (criminal proceedings) centred on a past event.

Indeed, what comes to the fore is that, given these premises, risk prevention, whether of endo-procedural or merely social origin, is traditionally incorporated within both precautionary and sentencing decisions; thus, what comes to the fore is that in both decision-making moments, the two concepts of repression and crime prevention inevitably merge.

In fact, the possibility of issuing a prognostic assessment and at the same time preventing the offence is located within precise choices entrusted to the public authority.

In the penal system, it should be noted that, in addition to the general-preventive matrix already mentioned, in special-preventive terms, the risk of future dangerous behaviour can be considered both an element to be assessed at sentencing (e.g. as an aggravating fact, as indicated by numerous sentencing guide-lines in common law countries), and at the same time, the basis for the application of security measures⁴³⁷.

⁴³⁶ *Ibidem*.

⁴³⁷ They are provided for by several European legal systems, such as in the German legal system (*Maßregeln der Besserung und Sicherung*), the Italian legal system and those of at least six other Council of Europe member states. In fact, what is worth clarifying immediately is that a substantial number of European countries (contrary to common law and, in particular, American common law jurisdictions) have adopted the 'double track', typical of the current Italian penal code, distinguishing between penalties and security measures, the former based on culpability, the latter based on the dangerousness of the individual. It seems important to point out that there are often significant inconsistencies, between jurisdictions, in the regulation of measures based on dangerousness. As the European Court of Human Rights pointed out in a well-known judgment, 'the same type of measure may be qualified as an additional sanction in one state and as a preventive measure in another'¹⁶. Regardless of the *nomina iuris*, the focus here is on the wide range of 'restrictive' measures that courts or other public authorities (such as probation boards, of Anglo-Saxon origin) may impose on persons who, having been charged with or convicted of an offence, are considered potential repeat offenders. It should be pointed out that these institutions, whatever their qualification, are distinct from another range of provisions, *preater delictum*, which in our legal system are properly called preventive measures and which fall outside the scope of this investigation. The focus, here, is on the measures by means of which systems intervene on the personal liberty of those accused or convicted of committing a crime, with the specific intention of circumscribing, for the future, their social dangerousness.

In other words, moreover, social dangerousness may even fall into all the categories already mentioned, aggravating, at the same time, the sanction imposed by the judge.

In conclusion, it is considered that, also due to issues mainly of focus on a specific aspect and therefore also of method, it is not possible to simultaneously consider the possibility of applying risk assessment in these two macro-areas, even though they both present profiles of marked and relevant interest in the current debate.

We would like to briefly mention a consideration that relates to the line of study that we decided to set aside: the introduction of such risk assessment tools in the jagged landscape of precautionary measures. Well, it should be noted that in fact, it is precisely in this field that this type of instrument could certainly find fertile ground for application. This is because it is no coincidence that they are normally applied in the initial phase of proceedings; it is precisely at this juncture that the competent authorities have very little information either on the facts that have occurred or on the personality of the suspect. Precisely these conditions represent two decisive factors in the aggravation of the 'predictive' task to which the judge is called in all respects. Therefore, the reason why such tools could favourably be applied is that risk assessment tools seem to offer great support to this type of decision-making, providing it with the data and tools it needs⁴³⁸.

However, as already anticipated, it has been decided to embark on a path that looks, from a theoretical and practical point of view, at the application of such tools in the very delicate phase of the decision on sanctioning treatment, which in itself is closely connected to (or rather has within it) an assessment of dangerousness.

3.2 A decision looking to the future: between limits and difficulties of prognostic evaluation

As has already been mentioned in the previous lines, the assessment of social dangerousness is as difficult as ever and has always been the subject of debate and in the crosshairs of criticism for its 'undefined' characteristics and for the partial and more evident subjectivity that falls within this type of assessment.

More specifically, the attitude of scepticism regarding such verifiability of subjective status on the part of the defendant is partly explained by the intrinsic characteristics of the same type

⁴³⁸ As already extensively emphasised, the venue in which this assessment takes place (investigation or trial) specifically characterises the judge's assessment process, because of the different objective that the restrictive measure has to achieve and the different support that psycho-criminological science can offer to the assessment of the social dangerousness or flight risk of the suspect/defendant. For this reason, the reflection on the impact of risk assessment tools must be differently articulated according to the trial moments concerned.

of assessment. In essence, the question arises as to whether or not the aspiration to recognise the 'probability' that a given person commits 'new acts provided for by law as offences', as provided for in Article 203 of the Criminal Code, is really realistic⁴³⁹.

It is indeed necessary to assess that a significant obstacle to the reliability of this type of assessment depends on its intrinsic nature as a prognosis, which undoubtedly differentiates it from factual ascertainment⁴⁴⁰.

4 *The judge's decision and the beyond reasonable doubt criterion in prognostic evaluations*

On closer inspection, when a judge finds himself having to make assessments that have prognostic considerations within them, for example in the case of the assessment of the possible and current dangerousness of the individual, it must be borne in mind that these assessments in themselves affect certain decisions concerning the most delicate aspects of the judicial decision.

For example, a judge of the supervisory court, when he is faced with the task of improving the re-education of the convicted person or preventing the possible danger of committing other offences, must take into account a plurality of normative and jurisprudential indices in formulating his prognostic judgement⁴⁴¹.

Undoubtedly, the evaluations on the future and the prognoses entrusted to the judging body run through the entire criminal system and system: from the choices of incrimination to the formulation of the precept, from recidivism to the system of penalties and security measures. It is no coincidence, in fact, that the delicate junction of the entire general theory of crime, prognostic judgments lie at the heart of the penalty system. Although it is indeed a punitive system inspired by the retributive ideal, prognostications represent almost an 'extraneous' element. In fact, one could be content to establish the proportionate measure of punishment to be inflicted on the offender in relation to the disvalue of the act committed and limit oneself to that. However, it is considered that the retribution could be recognised as having a function as a theoretical limit to the pursuit of the purposes of prevention (both special and general) that

⁴³⁹ Cfr. M. MONTAGNA, *I confini dell'indagine personologica nel processo penale*, Rome, 2013, 77 ss.; E. MUSCO, *Misure di sicurezza*, 767; M. PELISSERO, *Pericolosità sociale e doppio binario*, Turin, 2008, 110 ss. and 344 ss.

⁴⁴⁰ When it is the judge's task to establish whether the defendant has committed the act described in the indictment, he scrutinises the past and focuses his attention on what has already happened; on the contrary, the one concerning social dangerousness is a verification that is projected forward, towards future scenarios.

⁴⁴¹ In that case, a fortiori, it cannot "remain entirely anchored to whether or not the applicant has a job", thus on the point G. MAGLIOCCA, *Attività lavorativa e giudizio prognostico finalizzato alla concessione dell'affidamento in prova*, in *Processo penale e giustizia*, No. 6, 2017.

guide the prognostic judgement. It is believed that in order to achieve the purposes of prevention of punishment, both in the phase of abstract formulation of the cases and of the edictal frames⁴⁴², and in the subsequent phase (the one that interests us here) of the concrete application of the punitive treatment, it would be necessary 'a great effort of rationalisation and coordination on the part of science, legislation and the dogmatics of criminal prognosis'⁴⁴³.

However, it is precisely the finalistic orientation and the necessary individualisation of the punitive response that impose an irreplaceable and constantly evolving role on prognostic judgements. Therefore, reflection here will attempt to *focus* on the typology of this judgement, on which of the various phases and types of judgement is involved, and on the extent to which it is considered essential to assess the type of assessment and its weakness.

As can be seen from the very same provision under Article 203 of the Criminal Code, the very phrase 'when it is probable that he will commit new acts provided for by the law as offences' undoubtedly affirms that social dangerousness is considered as the probability that new offences will be committed in the future. Indeed, this criterion of ascertainment engages the judge in a prognostic judgement that is fraught with uncertainty since it presumes to predict the offender's future criminal behaviour.

Prognostic judgements play a fundamental and irreplaceable role in the penal system. It is no coincidence that the first applications of artificial intelligence tools in the criminal justice system concerned support in the assessment of recidivism, which is one of the fundamental tasks of the criminal justice system, which therefore passes through the formulation of prognoses on the future behaviour of the offender and on the preventive effectiveness of criminal sanctions⁴⁴⁴.

Therefore, as already anticipated, one of the protagonists of the prognosis in the penalty system is undoubtedly the judge, the one who can and does know the characteristics and needs of the offender (and at the same time assess the disvalue of the offence committed)⁴⁴⁵. And it is precisely the judge who, through his legislatively constrained discretionary power, is called upon to calibrate the finalism of the punitive response through prognostic judgments.

⁴⁴² On closer inspection, it is considered that it is not only the prognostic judgments made by the judge on the risk of reoffending and the effects produced by the punitive response that are of decisive importance for the preventive efficacy of punishment, but also the prognoses made on these same aspects by the legislature.

⁴⁴³ L. MONACO, *Prospettive dell'idea dello 'scopo' nella teoria della pena*, Naples, 1984, 121 s.

⁴⁴⁴ If, as has been pointed out by the Cambridge Study in Delinquent Development (CSDD), most offences are committed by a small proportion of offenders, then the penal protection of the fundamental interests of the community passes through the prognoses on recidivism and the special-preventive effects of the penalty system. Thus, on this point, On the data of the empirical study on recidivism, see G. ZARA – D. FARRINGTON, *Criminal recidivism: explanation prediction and prevention*, 48.

⁴⁴⁵ F. BRICOLA, *La discrezionalità nel diritto penale*, Milan, 1965, 118 ss.

Therefore, it is precisely the purposes of special prevention, whose gaze is necessarily turned to the future, that impose an assessment of the impact of the punitive response on the offender's future life. But indeed, it is also the purposes of general prevention that condition judicial prognosis. Indeed, the formulation of the prognosis, even if ideally analysed on the side of special prevention, is imbued with general-preventive considerations *per se*.

However, the judge cannot be considered the only actor acting in prognostic evaluations; indeed, according to another upstream perspective, following a purpose-oriented criminal law, the prediction of future events (and hence their prevention) is an integral part of the criminal policy programme formulated by the legislator. The important element, however, against which the judge's assessment 'clashes', which is often lacunar or incomplete, relates precisely to the criteria entrusted to him in order to make this type of assessment. Reference is made to all those cases in which the parameters possess vague contours or a polysense meaning that therefore affect the determinacy of the prognosis⁴⁴⁶.

In this context, for example, the glaring vagueness of the parameters and prerequisites that characterise the regulation of suspended sentences is undoubtedly in evidence, transforming the prognostic judgement into a prophecy⁴⁴⁷ in certain respects. Or again (and this will be the central theme of the following chapter⁴⁴⁸) the very concept of capacity to commit offences regulated in Article 133(2) of the Criminal Code is considered so ambiguous that it allows diametrically opposed interpretations.

Undoubtedly, on the opposite front, different considerations could be made with regard to those cases in which the legislature has set up a system in such a way as to leave the judge almost no room for discretion; in the case, for example, of the rigid prognoses on social dangerousness that have long governed security measures.

Undoubtedly, in conclusion, looking at the context in both directions, the relationship with scientific knowledge is extremely delicate: not only to support and validate the judge's prognoses, but also to check the empirical rationality of certain prognostic generalisations formulated by the legislator. There is no denying that the role of the legislator, who finds himself having to establish commensuration criteria and mechanisms regulating the execution of the penalty, is a very arduous one; even then, these types of choices imply in themselves a

⁴⁴⁶ These formulas are, in some cases, so vague as to raise the suspicion of constitutional illegitimacy due to violation of the principle of legality (Article 25(2) of the Constitution), as in the case of the notion of social dangerousness in Article 203 of the Criminal Code. Please refer on this subject to M. PELISSERO, *Pericolosità sociale e doppio binario. Vecchi e nuovi modelli di incapacitazione*, Turin, 2008, 115.

⁴⁴⁷ T. PADOVANI, *La disintegrazione attuale del sistema sanzionatorio*, in *Riv. it. dir. e proc. pen.*, 1992, 428.

⁴⁴⁸ Refer to Chapter 4.

formulation on the prognosis and future behaviour of the offender⁴⁴⁹. Indeed, within a punitive system, which is oriented towards special prevention purposes, the identification (in the abstract) of the punitive response and the fixing (in concrete) of the latter, have the ambition of directing human behaviour as effectively as possible.

In conclusion, in the vast and variegated universe of prognostic judgments, we intend to confine our analysis here to those within the penalty system, in particular those that concern positive special prevention: not only, indeed, the prediction of the offender's future conduct, but also the effects that the sanctioning response produces in terms of preventing the risk of re-offending. Here, therefore, we shall not deal with the general-preventive significance of prognostic judgements (unless it is functional to the treatment of prognosis oriented special prevention⁴⁵⁰). The study in this paper will focus, in fact, on understanding how predictive algorithms could be used as a support to the judge in the assessment of social dangerousness in the choice of punitive treatment made by the supervisory judge and as a support in the application of Article 133, paragraph 2 of the Criminal Code in the choice of punitive treatment.

It is necessary to make a final reflection linked to the fact that while on the one hand it is certainly true that the judge has the burden of establishing whether "it is probable that [the defendant] will commit new facts provided for by law as offences" - according to the provision under Article 203 of the Criminal Code, however, on the other hand, he certainly cannot confuse a reference - more or less explicit - to probability; indeed, if in order to pronounce sentence the defendant must be found 'guilty of the offence charged against him beyond all reasonable doubt, the same and equal criterion should be adopted for the verification of the probability of recidivism. Undoubtedly, this solution is imposed for an obvious reason: the attribution of the fact to the defendant implies a penalty; the declaration of social dangerousness implies instead a security measure. Therefore, given that in both cases it is personal liberty that

⁴⁴⁹ Indeed, it is considered that if the principle of the re-educative purpose of punishment, as the Constitutional Court has repeatedly reaffirmed, "constitutes one of the essential and general qualities that characterise punishment in its ontological content, and accompany it from when it comes into being, in the abstract normative provision, until when it is actually extinguished" ⁸, then every phase of punitive power, from the formulation of the sentence framework to its implementation, is marked by prognostic judgments on the future conduct of the offender. Thus, on this point Thus Constitutional Court, 21 September - 10 November 2016, no. 236; refer to the comment of the decision by F. VIGANÒ, *Un'importante pronuncia della Consulta sulla proporzionalità della pena*, in *DPC*, 14th November 2016.

⁴⁵⁰ By way of mere simplification, in fact, on the subject of suspended sentences we shall address the relations between the objective and subjective requirements for access to the benefit. The sentence limit identified by the legislator is, in fact, an expression of the general-preventive requirements which should not, however, guide the judge in the formulation of the prognosis. The prognostic judgement is, in fact, a different and additional requirement for the applicability of the suspension institute which should respond exclusively to special prevention requirements.

is called into question, there is no valid justification for considering that the rule of judgment should be more severe in one case and less so in the other.

Indeed, following the same line of reasoning, the defendant can and should be considered 'socially dangerous' only if it is probable, beyond reasonable doubt, that he will commit future offences; indeed, the probability of recidivism is what certainly needs to be demonstrated; whereas the criterion of reasonable doubt marks the quantum of proof required to achieve such a demonstration⁴⁵¹.

Moreover, applying such an apparently severe rule of judgement would allow, even in the absence of any regulatory changes⁴⁵², to circumscribe the operation of security measures to cases indicative of positive prognosis⁴⁵³.

5 *Prognostic assessments referred to the judge in the Italian legal system*

On closer inspection, as already mentioned, there are several cases in which the judge finds himself having to make prognostic evaluations: first of all, those on the choice of security and prevention measures, those referred to the supervisory courts in the phase of execution of the sentence and, finally, all the cases in which the judge must decide on the quantum and what sentence to impose on a subject (the sentencing phase which will be discussed in the next chapter). Indeed, one cannot disregard the many additional venues where the judge is asked to make a criminal prognosis. One thinks, for example, of the suspended sentence, or the discipline of probation⁴⁵⁴. An attempt will therefore be made, briefly, to illustrate the type of assessment envisaged for the cases mentioned in order to serve as a background and outline for the specific case to be dealt with in the following pages⁴⁵⁵.

⁴⁵¹ On the applicability of the 'beyond reasonable doubt' rule to contexts other than the decision on guilt; P. FERRUA, *La prova nel processo penale, I, Struttura e procedimento*, 2a ed., Turin, 2017, 92 ss.

⁴⁵² In the doctrinal debate, 'social dangerousness' is often portrayed as an institution in crisis: beyond the doubts as to the relevance of the label 'socially dangerous', i.e. as a likely perpetrator of future offences, the question arises as to whether it is really possible to rationally verify the risk of recidivism.

⁴⁵³ In this regard, one may recall the work of the 'Pelissero Commission', which - with the aim of imposing 'greater rigour on the judge in affirming the existence of this element' - intended to anchor the status of socially dangerous person to the 'relevant probability' of committing crimes. A strict observance of the rule of judgement in question may already enable such effects to be achieved.

⁴⁵⁴ This discipline of probation, introduced in 2014, which can only be ordered when the judge - 'on the basis of the parameters of Article 133 of the Criminal Code' - 'considers the treatment programme presented to be suitable and believes that the defendant will refrain from committing further offences'.

⁴⁵⁵ One cannot overlook the many additional venues in which a criminal prognosis is required. One thinks of the suspended sentence, or - to take a recent example in the procedural sphere - of the discipline of probation, introduced in 2014, which can only be ordered when the judge - moreover, 'based on the parameters set out in Article 133 of the Criminal Code' - 'considers the treatment programme presented to be suitable and believes that the defendant will refrain from committing further offences'.

The problem that immediately comes to the fore in criminal prognosis, i.e. the judgement aimed at understanding whether or not the offender will commit new offences, is characterised by its difficulty of definition and evanescence, since it is substantially based on intuitive parameters left to the judge's fair appreciation⁴⁵⁶. Indeed, the basis of the judge's assessment must be guided by the indices indicated in Article 133 of the Criminal Code⁴⁵⁷. which distinguishes criteria for verifying the seriousness of the offence and criteria for identifying the offender's capacity to commit offences. Indeed, it is precisely the latter that highlight the fragility of the concept of social dangerousness, since they shift the focus from the fact to the subject, assessing his responsibility not so much for what he has committed, but more so for his own conduct in life⁴⁵⁸.

On closer inspection, the coexistence in the Italian legal system of security and prevention measures opens up quite a few interpretative questions, arousing considerable perplexity. Both undoubtedly converge in the purpose of "neutralising" social dangerousness and diverge in their relationship with the criminal judgement: in fact, on the one hand, security measures presuppose the judicial ascertainment of the crime and are therefore post-delictum sanctioning measures; on the other hand, prevention measures disregard the ascertainment of the crime committed, configuring themselves instead as *praeter delictum* measures. However, there is the common foundation and common purpose that both measures possess⁴⁵⁹.

It should be pointed out from the outset that in today's doctrinal debate the concept of 'social dangerousness' is often represented as an institution in crisis: indeed, beyond the doubts as to the topicality of the label of 'socially dangerous' subject, as a probable perpetrator of future

⁴⁵⁶ E. MUSCO, *La misura di sicurezza detentiva. Profili storici e costituzionali*, Milan, 1978, 191 ss.; M. BERTOLINO, *Profili vecchi e nuovi dell'imputabilità penale e della sua crisi*, in *Riv. It. Dir. Proc. Pen.*, 1988, 252; P. NUVOLONE, (voce) *Misure di prevenzione e Misure di sicurezza*, in *Enc. dir.*, XXVI, Milan, 1976, 631; E. MUSCO, (voce) *Misure di sicurezza*, in *Enc. Dir.*, Agg. I, Milan, 1997, 762; C. PELUSO, (voce) *Misure di sicurezza*, in *Dig. Disc. Pen.*, VIII, Turin, 1994, 145.

⁴⁵⁷ One has to deal today with a fallible provision such as that provided for in Article 133 of the Criminal Code. It is, in fact, a The consequence of this "inadequate procedural instrumentation" is twofold: "on the one hand, the tendency to blur the ascertainment of the character and personality, in general, of the offender into the realm of intuitions and impressions of atmosphere, which, in addition to being difficult to translate into writing, are not easily verifiable; on the other hand, a certain inertia on the part of the judges, both in motivating and in dealing with the investigation of the second part of Article 133 of the criminal code". Thus F. BRICOLA, *La discrezionalità*, 116.

⁴⁵⁸ U. FORNARI, *Trattato di psichiatria forense*, 55, the author states that: "*Lombroso wished to shift the study of crime from the fact to the individual who committed it, elaborating a bio-anthropological doctrine that was de-emphasising, deterministic and reductive, and which received and still receives so much fame throughout the world. It arose from a chance observation about the existence, where there should have been a bone ridge, of a congenital morphological anomaly in the skull of a criminal. This man's skull conventionally constitutes the birthplace of criminal anthropology. Thus arose the stereotype of the born criminal*".

⁴⁵⁹ See on this point the decision of the Constitutional Court No. 68 of 1964; the Court, in Judgment No. 177 of 1980, reiterated the opinion, noting that these are two species of the same *genus*.

offences, one wonders whether it is actually possible to verify, passing through the 'meshes of rationality', the risk of recidivism⁴⁶⁰.

At this stage, therefore, an attempt will be made to trace the current coordinates of the prognostic assessment of 'social dangerousness', trying to analyse the salient features of its ascertainment. Finally, some *de iure condendo* reflections will be sketched out for a discipline in which prognostic tools, such as predictive algorithms, could usefully see the light of day.

In attempting to outline and provide an analysis of the role of prognostic judgments in the penalty system, it appears useful to start, from the outset, from the normative datum. In this regard, in fact, it must be emphasised that the expressions used by the legislature are varied and disparate. Indeed, prognoses operate both in the choice and commensuration of the penalty, sometimes anticipating the ascertainment of the fact, and in the execution phases of the same. In this regard, in order to analyse the role of prognostic judgments in the current Italian penalty system, it would seem appropriate to follow a scan that is itself dictated by the timing of the criminal proceedings⁴⁶¹.

5.1 *The different types of prognosis at the trial stages*

As anticipated in the preceding paragraph, within the different procedural phases there are time scans in which the judge has to operate and make a prognostic judgement.

Just to outline the fundamental lines and scans, it should be noted that a first phase is devoted to prognostic judgments that are formulated before the ascertainment of the fact. In particular, these are assessments of the future conduct of the offender and which, by their nature, are characterised by a tendency to incompleteness and lacunae of the cognitive elements available to the judge to decide. An example already mentioned here is the case of proceedings with probation pursuant to Article 168 *bis* of the criminal code.

A second part and second phase concerns instead the prognosis made by the judge of cognition at the end of the fact finding. One thinks, for example, of the assessment of criminal capacity in the commensuration of the sentence and the suspended sentence. In such cases, the judge has more information about the fact and the offender, but his prognostic judgment is constantly vitiated by a lack of the knowledge needed to formulate it. This is because not only is there a ban on criminological expertise at this stage, pursuant to Article 220 of the Code of Criminal Procedure, which severely undermines, for various reasons, the reliability of the

⁴⁶⁰ A. CABIALE, *L'accertamento giudiziale della pericolosità tra presente e futuro*, in *Arch. Pen.*, No, 2, 2022.

⁴⁶¹ S. ROMANO, *Il ruolo delle prognosi nel sistema sanzionatorio*, University of Milan 2018.

prognosis, but also because there is a constant cultural attitude of mistrust or underestimation of the importance of prognosis in the penalty system.

Lastly, a third stage concerns those prognostic judgements that are of decisive importance in relation to numerous alternative responses to custodial sentences at the execution stage. In this regard, one thinks of alternative measures to detention. Indeed, it is precisely the sources and tools that offer the judge, also through the work of the prison operators, a potentially complete cognitive picture of the offender and of the predictive factors of recidivism in the case in question that are the subject of in-depth study. Here too, in fact, the prognostic assessment of the risk of reoffending is carried out on the basis of maxims of experience and the personal intuition of the person called upon to decide, without sufficient recourse to empirical and scientific knowledge.

Indeed, the prognostic activity begins even before, in many cases, the pronouncement of conviction and the prognostic judgements themselves play a very delicate and at the same time fundamental role in the (possible) precautionary phase.

In conclusion, today, the *world of punishment* has been progressively invested with the tasks of social defence and the management and treatment of the risk of reoffending. And it is in this area that the legislator has, in recent years, made increasing use of prognostic assessments.

5.2 *Prognostic evaluation in security measures*

The concept of dangerousness can be analysed with reference to the potential dangerousness of a subject⁴⁶², i.e. irrespective of the assessment of the concrete manifestation of this personal quality through the actual perpetration of a certain offence or other dangerous human act; at the same time, it can also be understood as an assessment of a prognostic nature; in particular, of a judgement that is implemented post-offence, projecting into the future, with the aim of verifying the degree of probability that the person may return to committing criminal offences again. It is precisely in this context that the application of security measures can be placed⁴⁶³.

⁴⁶² Jurisprudence has also expressed itself on the subject, going so far as to affirm that the jurisprudence declines the concept in the following terms: "Dangerousness is a quality, a way of being of the subject, from which the probability that he commits new crimes is deduced. it differs from criminal capacity, which always exists to a more or less accentuated extent, for the very fact that the subject has already committed the crime and therefore constitutes a subjective aptitude to the commission of the crimes themselves. Criminal capacity is therefore the genus and dangerousness the species, since the former is only a possibility, while the latter is the probability of committing criminal offences. dangerousness coincides only with the prognostic - preventive dimension of criminal capacity but not with its ethical - retributive dimension". (Cass.,II, 5.6.1990,n.9572 ,Aresu,CED).

⁴⁶³ "In relation to the nature of the feared damage, or rather to the manifestations through which the probable damage may arise, many Authors, with specific reference to the post-delinquent and extra-delinquent dangerousness, instead of speaking of social dangerousness prefer the expression "criminal dangerousness", since it is the probability of the commission of future crimes that is under discussion. The expression 'social

On closer inspection, in the case of security measures, prognosis has undoubtedly been the subject of wide-ranging and in-depth examination and treatment; this interest is explained not only because the ascertainment of social dangerousness constitutes one of the application prerequisites of security measures, but, even before that, because the prevention of recidivism, through care and control, is the very essence of the security binary of the Italian penal system. In fact, to date, even the world of punishment has been progressively more and more invested with the tasks of social defence and management of the treatment of the risk of reoffending. It is precisely for this reason that the legislator has made increasing use of prognostic assessments in this area⁴⁶⁴. It should be noted from the outset that special prevention and prognosis have chosen the system of security measures (and not punishment⁴⁶⁵) as their preferred field of action.

5.3 *Prognostic evaluation in prevention measures*

Also, in the context and in the sector of the measures of prevention, since a determinate quality of the person is under discussion and not the single concrete facts, the judgement of social dangerousness maintains in itself a so-called judgement of "probability"⁴⁶⁶. Clearly, any kind of "certainty" is excluded from this type of evaluation; it is, for the most part, an "evaluation of an essentially symptomatic character"⁴⁶⁷. In fact, it is mostly a matter of 'statements of probability as to the future legal conduct of the subjects' which, being resolved in a judgement with a probabilistic structure, can only be made by evaluation of the factual

dangerousness' actually has its *raison d'être* in the code, in the circumstance that, since the security measures also apply to persons who cannot be charged, it was considered that the expression itself better corresponded to what the aforementioned persons may fear, that is, not a punishable offence, but only an act that is foreseen as a crime", thus B. SICLARI, *Applicazione ed esecuzione delle misure di sicurezza personali*, Milan, 1977, 17.

⁴⁶⁴ Thus, on this point, "The legislator of the reforms" while not providing any notion of qualified dangerousness "expands the scope of its operation, conditioning the granting of new institutions in the function of special prevention to negative prognosis of recidivism". One thinks of the substitute sanctions pursuant to Article 58 of Law 689/81; of the alternative measures pursuant to Article 47 et seq. of the Criminal Code; of the conditional suspension of sentences pursuant to Article 164 of the Criminal Code and, most recently, of Article 4 of Law 67/2014, trial suspension with trial); M. BERTOLINO, *Declinazioni attuali della pericolosità sociale: pene e misure di sicurezza a confronto*, in *Arch. Pen.*, 2014, 461.

⁴⁶⁵ Refer to Chapter 4.

⁴⁶⁶ P. MILETO, *Le misure di prevenzione*, in G. Ambrosini - P. Miletto (eds), *Le sostanze stupefacenti. Le misure di prevenzione*, in *Giurisprudenza sistematica di diritto penale*, F. Bricola - V. Zagrebelsky (eds), Turin, 1989, 145 - 146.

⁴⁶⁷ Cass., 14 February 1997, Nobile and others, in Cass. pen. 1997, 3171; Cass., 19 December 1996, Di Muro, in Cass. pen. 1997, 2576; S. P. FRAGOLA, *Le misure di prevenzione*, Padua 1992, 14; on different methods of criminal prognosis, namely the intuitive method, of debated reliability, and the statistical method, which uses negative values collected in prognostic tables as symptomatic indices of dangerousness, which, although it has a greater scientific foundation, has not yet been sufficiently developed through empirical research on all criminal groups or groups with an expressed dangerousness; in fact, it is reported that the intuitive method is used in practice and that scientific research in the sector is not sufficiently committed.

elements and, at the same time, by approximation. From this type of judgement, the precariousness of judgements based instead on prognosis is, in part, excluded⁴⁶⁸.

On closer inspection, recent interpretative frameworks that have emerged at the national and even at the supranational level, both on the jurisdictional side and on the regulatory planning side, lead one to reflect on an issue free from preconceptions on the subject of social dangerousness; in particular on the question pertaining to its 'systematic citizenship' and its sustainable and controversial contexts of ascertainment. Since its inception, in fact, it has been criticised and considered controversial due to its ill-defined and descriptive contours. As already mentioned above, therefore, also in the context of preventive measures, social dangerousness accounts for a substantial part of the current regulatory framework in the criminal and para-criminal field.

6 *The structural characteristics of the prognosis of dangerousness*

On closer inspection, as has already been mentioned in part, the difficulties and criticisms levelled against the extremely arduous assessment and ascertainment of social dangerousness depend, to a large extent, on scepticism as to the verifiability of this subjective condition in the defendant. Basically, what is most in evidence is the question of whether the aspiration to recognise the 'probability' that a given subject will commit new crimes (as provided for by Article 203 of the Criminal Code)⁴⁶⁹ is really realistic. Indeed, a significant obstacle to the reliability of such an assessment depends first and foremost on its nature as a prognosis, which undoubtedly differentiates it from the assessment made merely on the fact⁴⁷⁰. When, in fact, it is up to the judge to establish whether the defendant has committed the act described in the indictment, the judge finds himself having to make a double assessment: he must at first look at the past and focus his assessment, on the basis of the evidence, on what has already happened; on the other hand, the analysis and assessment of the social dangerousness of an individual is a very different verification since it is an assessment that is projected forward and looks to the future⁴⁷¹.

⁴⁶⁸ W. HASSEMER, *Einführung in die Grundlagen des Strafrechts*, München 1981, 244 ss. See also A. MANNA, *Il diritto delle misure di prevenzione*, 14 ss.

⁴⁶⁹ M. MONTAGNA, *I confini dell'indagine personologica nel processo penale* 77 ss.; E. MUSCO, *Misure di sicurezza*, 767; M. PELISSERO, *Pericolosità sociale e doppio binario*, 110 ss.

⁴⁷⁰ In this sense F. CAPRIOLI, *Pericolosità sociale e processo penale*, 23 ss.; T. PADOVANI, *Fatto e pericolosità*, 121; F. TAGLIARINI, *Enciclopedia del diritto*, Vol. XXXIII, voce *Pericolosità*, Milan, 1983, 29.

⁴⁷¹ One element that undoubtedly accumulates the two types of situations in which the adjudicating body finds itself is that it is equally true, however, that in both cases the judge has to deal with circumstances of which he has no direct knowledge: of course, he cannot witness events that have not yet taken place, but neither has he perceived the past events that are the main subject of the proceedings. From this point of view, therefore, the

Indeed, the ascertainment of the main fact has some important advantages: firstly, it concerns the types of evidence potentially available to verify the validity of the charge. Indeed, it should be noted that, precisely because the object of ascertainment in such a case is the past, it is possible that it may have somehow been captured (for instance, in the case where the fact has been perceived by several reliable witnesses, or where there are some recordings); in such a case, the traces of the past are solid and therefore the judge is unlikely to make a mistake in affirming the innocence or guilt of the defendant. On the other hand, the case is different when it is an assessment linked to a prognosis: future events cannot be 'reproduced' or shown at trial. The only tool the judge has consists of reasoning and reflection that starts from the past and the present and tries to identify and glimpse the signs of a (only) possible future. Indeed, it can also be said that the assessment of social dangerousness is always based on 'circumstantial' evidence that can never directly represent what is to be proved⁴⁷².

Moreover, a further profile that makes the decision on the risk of reoffending even more fragile is undoubtedly linked to the unpredictability of future scenarios: while indeed on the one hand the past is unchangeable, on the other hand the future could be affected by completely and utterly unknown factors; consequently, a correct interpretation of the 'evidence' concerning the defendant's subsequent conduct could also be contradicted by unexpected circumstances that interrupt the previously outlined causal chain. Likewise, however, precisely because it is an assessment of the future, such a prognosis enjoys a certain degree of empirical verifiability. Consequently, a 'false positive' could be recognised in the course of the treatment observation and thus lead to the interruption or mitigation of the measure applied⁴⁷³.

In conclusion, therefore, the judgement on the fact and the judgement on dangerousness may be characterised by certain common cognitive obstacles: firstly, the circumstance that one is dealing with facts that the judge has not perceived or cannot perceive directly; secondly, also the absence of evidence 'representative' of what one wants to prove.

It is relevant to consider that an exclusivegnoseological limitation of prognostic judgement is, on the other hand, precisely the inescapable rate of unpredictability of the evolution of the

findings in question suffer from the same problem, since it is in any case with the unknown that one has to deal. See, M. TARUFFO, *Il giudizio prognostico del giudice tra scienza privata e prova scientifica*, in ID., *Sui confini. Scritti sulla giustizia civile*, Bologna, 2002, 335, who states that, in any case, 'the judge is dealing with a hypothetical statement of fact'.

⁴⁷² Againg, M. TARUFFO, *Il giudizio prognostico del giudice tra scienza privata e prova scientifica* 335 s.

⁴⁷³ See M. PELISSERO, *Pericolosità sociale e doppio binario*, who points out that 'the false negative can be disproved by the commission of the offence and could thus justify the introduction of new security measures [...]; on the other hand, on the other hand, once the measure has been introduced into the system, it would be very difficult to prove its uselessness, since it is not possible to ascertain the false positive, i.e. that the positive prognosis of dangerousness is false with respect to internees'.

present to which one can only subsequently adapt, at least to a certain extent. These considerations and elements described were undoubtedly taken into account by the legislature. Initially, indeed, the check on social dangerousness is carried out both at the cognitive and enforcement stages. This further check, provided for in Article 679 of the Code of Criminal Procedure, is fundamental not only to understand whether the dangerousness may have 'diminished' in the meantime, but may at the same time act as a guide to reorient previous assessments. In addition, the status as a 'socially dangerous' person should be periodically reviewed and re-evaluated - pursuant to Article 208 of the Criminal Code - and the measure could also be revoked in advance, pursuant to Article 69(4) of the Criminal Code.

Just like the assessment pursuant to Article 679 of the Code of Criminal Procedure, re-examination and revocation, echoing institutes of the precautionary procedure, serve both to adjust the conditions of personal liberty to the treatment response and to remedy any shortcomings of the prior checks. None of this has been provided for in relation to the ascertainment of the accused fact; the only possibility of review is anchored to the strict rules of revision (Article 629 et seq. of the Code of Criminal Procedure).

Clearly, the first question and the starting point from which it was decided to start the development of this work starts from a general question and from the positive answer of the same comes the reflection that concerned and approached, at the same time, the instruments of I.A. and this type of assessment.

Therefore, the first point of departure has more to do with a choice of legislative policy than with a dogmatic notion: in fact, one wonders whether it is possible to think of doing without, and therefore of renouncing, the normative notion of social dangerousness, understood as the "subjective - albeit transitory - qualification of an individual and deriving from a twofold assessment operation", that of appreciating past conduct, from which a reliable prognostic judgement on future conduct takes its cue⁴⁷⁴? Clearly this is a rather delicate issue and question. Indeed, it relates to and intersects with the issue of regulating the cognitive basis of such a judgement, which would be the only way to comply with constitutional provisions and repudiate an 'unacceptable subjectivism of assessment'⁴⁷⁵.

Although this category has been much criticised to date and presents profiles that collide with the guaranteed system of the criminal justice system, it is, unfortunately, a category that is currently indispensable from a legislative point of view.

⁴⁷⁴ See, R. MAGI, *Per uno statuto unitario dell'apprezzamento della pericolosità sociale. Le misure di prevenzione a metà del guado?*, in *DPC*, No. 3, 2017, 138 ss.

⁴⁷⁵ On this point, R. MAGI, *Per uno statuto unitario dell'apprezzamento della pericolosità sociale*, 138.

In fact, a total abandonment of the albeit controversial legal category would appear to be difficult to achieve in the current historical moment, however difficult it would then be to realise a model of constitutional and conventional compatibility, between its ascertainment and the different types of consequences on the one hand and, on the other, the values of effectiveness of the individual's rights.

7 *How Artificial Intelligence intervenes in the judgement of dangerousness*

On closer inspection, it was considered necessary to analyse the nature, form and evolution of the concept of social dangerousness on the one hand to take cognisance of the fact that it is a type of assessment that has always recurred in the criminal justice system (not only in Italy) and to see how it has evolved over time. But above all, what came to the fore is the fact that in itself, a judgement of this kind, entrusted to the judge, is a judgement totally steeped in the future, in prognostic judgements that are entrusted to a human being who finds himself making a judgement, often with few elements at his disposal.

What is certainly relevant is that, in recognising that the judgement of dangerousness per se implies the need for expert intervention, it is always the judge who has to issue the last word and thus his judgement. Indeed, even if the idea of scientific predictions of 'future criminality' has been abandoned, it is nevertheless possible to obtain useful data: 'a risk of psychopathological decompensation with concomitant, probable violent acts' would indeed be appreciable; in essence, the expert, by combining certain risk factors, might be able to provide the judge with a sort of risk assessment, at least as regards these types of behaviour⁴⁷⁶.

In fact, and what we propose in this paper, there are tools (as already anticipated in the previous pages), such as algorithms, which could be used for the assessment and evaluation of social dangerousness, based precisely on a prognostic evaluation of the probability that the subject returns to crime⁴⁷⁷ (or, for the purposes of preventive measures) delinquency tout court, both in the application of security measures (*ex art. 202 c.p.*) or in the application of prevention measures (Article 274, lett. c.c.p.), or in the application of suspended sentences (Article 164,

⁴⁷⁶ See A. CAPUTO, *La pericolosità sociale. Vecchie esigenze e nuove prospettive alla luce della legge 30 maggio 2014, n. 81*, Rome, 2015, 126 ss.; R. CATANESI - F. CARABELLESE-I. GRATTAGLIANO, *Cura e controllo. Come cambia la pericolosità sociale psichiatrica*, in *Journal of psychopathology*, No 1, 2009, 69, who point out that 'several risk factors have been identified that, in various combinations, are capable of significantly increasing the likelihood of new violent acts'; M. T. COLLICA, *Il giudizio di imputabilità tra complessità fenomenica ed esigenze di rigore scientifico*, in *Iris*, 2008, 1212; H. G. KENNEDY - F. CARABELLESE - F. CARABELLESE, *Evaluation and management of violence risk for forensic patients: is it a necessary practice in Italy*, in *Journal of Psychopathology*, 2021, 11 ss.; M. PELISSERO, *Pericolosità sociale e doppio binario*, 120; S. QUATTROCCOLO, *Artificial Intelligence*, 153 ss.

⁴⁷⁷ A. M. MAUGERI, *L'uso di algoritmi predittivi per accertare la pericolosità sociale*, in *Arch. Pen.*, No. 1, 2021.

paragraph 1 c.p, or of alternative measures to detention, at the time of execution and, finally, at the time of commensuration of the sentence, with regard to the element of capacity to commit offences, pursuant to Article 133, paragraph 2⁴⁷⁸ of the Criminal Code or in other cases⁴⁷⁹.

On closer inspection, new technologies, and in particular artificial intelligence systems, offer the advantage of being able to draw on and process immense amounts of data from sources such as case-law and legislative databases and collections of precedents and, through the use of highly sophisticated systems and devices, should make it possible to bring out relationships, coincidences and correlations that enable a person to be profiled in order to predict subsequent behaviour, even of criminal relevance.

Considering that an algorithm, as already exposed in the previous chapters, consists of a sequence of instructions that must necessarily be followed to transform an input into output⁴⁸⁰, these are tools that analyse "a very large number of data related to the past, being able to identify recurrences (elements that repeat themselves, also called patterns) characterised by a much more solid statistical basis (in some respects) than that emitted by human judgements⁴⁸¹. Clearly, within this group of proposed tools, there are also machine learning tools that learn from the past and then emit useful data in the future⁴⁸². Obviously, learning, in this case, is aimed at prediction, at solving cases other than those analysed that can then be used in the future⁴⁸³.

The reason why it is considered useful and reasonable to propose the entry of such tools (albeit under certain conditions that will be set out below), lies in the fact that it is believed that they could act with a 'more objective' and unbiased methodology compared to a prognostic assessment of this kind issued by human beings.

⁴⁷⁸ Refer to Chapter 4.

⁴⁷⁹ For example, the use of algorithms is also extended to the choice of the type of rehabilitation programme in certain areas such as that of sex offenders, which has been based in many jurisdictions, first and foremost in Canada, on the assessment of the risk of reoffending with systems based on Risk-Need-Responsivity⁵ (RNR), treatment must be proportional to the risk of committing a new offence), which use actuarial tools, based on empirically validated risk factors, drawn from the personal and criminal history of the subject (STATIC 99R, STABLE 2007, ACUTE).

⁴⁸⁰ E. ALPAYDIN, *Introduction to Machine Learning*, Cambridge, 2010, 1.

⁴⁸¹ M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, 10.

⁴⁸² P. DOMINGOS, *L'algoritmo definitivo. La macchina che impara da sola e il futuro del nostro mondo*, Bollati Boringhieri, 2016, Turin, 86.

⁴⁸³ This process is called generalisation: given a set of initial information, a rule must be extrapolated that is suitable for predicting and solving future cases that have not yet been analysed¹⁷: machine learning aims to predict a certain outcome. According to a well-known definition, "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E". See, T. MITCHELL, *Machine Learning*, McGraw Hill, 1997, 2; *amplius* P. FABBRI, *Cos'è l'intelligenza artificiale e quali sono le applicazioni attuali e future*, in www.zerounoweb.it, 2019.

Indeed, the advantage of using predictive algorithms could undoubtedly be represented by greater legal certainty; whereas up to now, the assessment of social dangerousness, especially in the prognostic part of the judgement (predictive with respect to the future) has mostly consisted of institutional judgements, based on the judge's personal experience and their common sense⁴⁸⁴. Indeed, as has already been anticipated, the judgement of social dangerousness for the purposes of applying security measures has serious and undoubted limits of scientific reliability of criminogenetic and predictive judgements, such that, as has been observed, it can 'provide pseudo-scientific bases to those forms of discrimination - social, political, cultural, religious, racial - so frequent in the twilight of democracies'.

Therefore, it is precisely the concept of dangerousness that is configured as a 'hybrid', connoted simultaneously by medical and also legal parameters, which are highly ambiguous and scientifically inconsistent⁴⁸⁵.

Indeed, in the assessment of social dangerousness in the Italian legal system, it has already been pointed out that Article 203 of the Criminal Code refers in paragraph 2 to the inescapable declination that has been made in the provision under Article 133 of the Criminal Code is like any prognostic judgement, based on the appreciation of the recurrence of a 'danger, which is necessarily and by its very nature directed to the future. This therefore excludes its possible declination in terms of historical certainty with an 'eliminable margin of fallibility'⁴⁸⁶.

⁴⁸⁴ G. CONTISSA-G. LASAGNI-G. SARTOR, *Quando a decidere in materia penale sono (anche) algoritmi e IA*: 631; C. BURCHARD, *L'intelligenza artificiale come fine del diritto penale? Sulla trasformazione algoritmica della società*, 1926, who highlights the shift of trust from people to technology. On the problematic nature of the prognostic assessment of dangerousness.

⁴⁸⁵ A. SALVATI, *La pericolosità sociale nell'ordinamento giuridico italiano*, in *Amministrazione in cammino*, 11th May, 2011; V. M. MASTRONARDI, *Manuale per operatori criminologici e psicopatologi forensi*, Milan, 1996, 388; on the problematic nature of the prognosis of social dangerousness, among others, M. BERTOLINO, *Il "crimine" della pericolosità sociale: riflessioni da una riforma in corso*, in *Riv. it. med. leg.* 2016, 1371 ss.; G. FIANDACA, *L'imputabilità nell'interazione tra epistemologia scientifica ed epistemologia giudiziaria*, in *Leg. pen.* 2006, 263; A. MANNA, *Imputabilità tra prevenzione generale e principio di colpevolezza*, in *Leg. pen.*, 2006, 241; ID., *Imputabilità, pericolosità e misure di sicurezza: verso quale riforma?*, in *Riv. it. dir. proc. pen.*, 1994, 1426 ss.; M. T. COLLICA, *La crisi del concetto di autore non imputabile "pericoloso" del reato*, in *Dir. pen. cont.*, 2012, 274 ss.; T. PADOVANI, *La pericolosità sociale sotto il profilo giuridico*, in *Trattato di criminologia*, F. Ferracuti (ed), vol. XIII, 318 ss.; D. PETRINI, *La prevenzione inutile*, Naples, 1996, 294 s.; A. MARTINI, *Essere pericolosi. Giudizi soggettivi e misure personali*, 155.

⁴⁸⁶ The Court (Court of Cassation, Sec. II, 11 August 2020, in Mass. Uff., no. 23797) goes on correctly pointing out that such feasibility "is all the more duly avoided the more the cognitive presupposition is strengthened, i.e. the analysis of everything that has emerged up to the time when the prognosis is required (modalities of the facts already realised, causal antecedents, antecedent life conduct, factors that may have affected the determination to act, ability to resist the urges that move towards the transgression of the precept)". F. BRICOLA, *La discrezionalità nel diritto penale*, Milan, 1965, 402; I. MERZAGORA BETSOS, *Imputabilità e pericolosità sociale: un punto di vista criminologico e psicopatologico forense*, in *Verso un codice penale modello per l'Europa. Imputabilità e misure di sicurezza*, di Manna, Padua 2002, 112; D. NOTARO, *Art. 203 c.p.*, in T. Padovani (ed), *Codice penale*, Milan, 2019, 1409 s.

8 *The problem of defusing human cognitive bias: possible advantages in the use of predictive algorithms*

Here we propose to propose a recourse to and thus a rapprochement between criminal justice and mathematical sciences; we proposed to approach this topic by looking at such sciences with fascination and prudence, so that the recourse to such science could be taken as functional to the objective of reducing the *bias* of retrospective judgement, predicting judicial errors.

As already anticipated, this desirable result could be achieved through the recourse to specific EISs that - relying on pre-set and automated calculation methods - could to all intents and purposes allow to defuse the human 'cognitive bias', induced, in the posthumous prognosis, by the knowledge that has arisen and the data available *ex post*.

Undoubtedly, one of the positive aspects of predictive algorithms would be that they 'design a normative procedure that moves from a set of data towards a desired output, excluding subjective intuitions and the arbitrariness of the process'. In this way, it represents a mathematical model that can be handled by a human being, even in criminal proceedings, provided that it is based on a validated theory and that this theory, at the same time, has been correctly encoded in the algorithm⁴⁸⁷.

Indeed, it is precisely by following this direction that it would be possible to develop the tendency that emerges in the most recent assessments of an actuarial nature, which constitute the theoretical prerequisite for the use of predictive algorithms, to affirm 'an evidence-based conception of the assessment of the individual risk of commission of a (new) crime: a conception, therefore, that remains based on objective evidence, which is then, at a later date, destined to supplant or supplement the judge's intuitive assessments that, to date, are still widely diffused⁴⁸⁸.

On closer inspection, one would value, for example, also in the prognostic phase, the objective and rational character of the assessment of social dangerousness, enhancing the data emerging in the cognitive phase of the judgement, based, above all, on the most recent

⁴⁸⁷ Indeed, these two requirements are crucial: [...], the possibility to review, discuss, challenge algorithms is a basic condition for fair criminal proceedings, in accordance with fundamental human rights'. Thus, on the point, S. QUATTROCOLO, *Intelligenza artificiale e giustizia*, 6.

⁴⁸⁸ F. BASILE., *Intelligenza artificiale e diritto penale*, 17; G. Zara, *Tra il probabile e il certo*, with particular reference to the work of J. P. SINGH (et oths)., *A comparative study of violence risk assessment tools: a systematic review and meta-regression analysis of 8 studies involving 25980 participants*, in *Clin Psychol Rev*, 31, 2011, 499 ss.

jurisprudence in the application of the measures of prevention on the ascertainment and reconstruction of facts or even of real judicial precedents⁴⁸⁹.

On closer inspection, the reflection that has been generated in the North American legal system is based for the most part on the exclusive reliance on the instincts and experience of the judge or of the person who has to decide must be modified at the present time because it is no longer sufficient or, worse still, may even be unethical since it could give rise to a sort of 'sentencing malpractice' that produces recommendations and sentences that are neither transparent nor entirely rational⁴⁹⁰.

Rather, the idea would be to select the possible options in the choice of sanction in order to identify the instrument that would best reduce the possibility of recidivism through re-education, incapacitation or deterrence; indeed, this would be a scientific question that should be best guided by the science of best practices, the so-called 'evidence-based practice'⁴⁹¹.

8.1 *The second step in risk assessment tools: the individual and the group*

On closer inspection, a fundamental step in the understanding of risk assessment is that inherent in the relationship between the group and the individual.

This is because the social sciences collect statistical data in order to arrive at general conclusions that are applicable to groups of cases⁴⁹².

Generally speaking, it can be seen that the accuracy of a given scientific prediction is based on the identification of a more or less large reference sample. If one applies this premise to risk

⁴⁸⁹ A. M. MAUGERI, *I destinatari delle misure di prevenzione tra irrazionali scelte criminogene e il principio di proporzione*, in *Indice Penale*, 2017, 37 ss.; A. M. MAUGERI – P. PINTO DE ALBUQUERQUE, *La confisca di prevenzione nella tutela costituzionale multilivello: tra istanze di tassatività e ragionevolezza, se ne afferma la natura ripristinatoria (Corte Cost. n. 24/2019)*, in *DPC*, 3, 2019, 97 ss.

⁴⁹⁰ Judges cannot impose appropriate sentences – those that will best help to protect public safety – without professionals conducting appropriate [evidence-based] violence risk assessment”. HART, *Evidence-Based Assessment of Risk for Sexual Violence*, 1 *Chapman J. Crim. Just.*, 2009, 143, 144. It is irrational and unethical to make predictions based on legal or clinical experience alone, because research has shown this to be an invalid and inaccurate method for assessing risk. Cfr. D. FAUST – J. ZISKIN, *The Expert Witness*, in *Psychology and Psychiatry*, 241 *Science* 31, 33, 1988, in particular “If expertise is defined solely by accuracy, the actuarial method is the ‘expert’”.

⁴⁹¹ “Thus, the concept of evidence-based practice in corrections refers to corrections practices that have been proven through scientific corrections research ‘to work,’ to reduce offender recidivism”. Indeed, they are defined as: ‘professional practices that are supported by the “best research evidence”, consisting of scientific results on intervention strategies derived from clinically relevant research. based on systematic reviews, reasonable effect sizes, statistical and clinical significance, and a body of supporting evidence.

⁴⁹² For instance, the 'G2I' (Group to Individual) paradigm is at the centre of the debate on forensic predictive tools. In fact, the prediction of future behaviour can only be based on the observation of a reference group and is a function that depends on the results of the group. Zara and Farrington, in their seminal work, report an iconic passage from the pen di Sir Conan Doyle, in which the author conveys precisely this concept, through the voice of Sherlock Holmes: “you can never foretell what any one man will do, but you can say with precision what an average number will add up to. Individuals vary, but percentages remain constant”.

assessment tools, one sees how a predictive assessment is only reliable if it correctly demonstrates the individual's propensity to reoffend by successfully applying the paradigm observed in the reference group to the individual.

In particular, the scale on which these two assessments operate is different and, for this reason, risk assessment tools are often criticised for the inaccuracy in the translation at the individual level of phenomena that are instead observed at the group level⁴⁹³.

8.2 *The advantages of a mixed algorithmic evaluation: the US example*

At a closer look, from the analysis of the functioning of the predictive software applied today in the American judicial system, it can be seen that it is based on an algorithm that analyses and processes different types of factors: first of all, what is undoubtedly relevant is that it processes and compares factors relating to data concerning the profiles of already convicted persons considered to be 'similar' to those of the defendant in the various cases under examination; the peculiarity lies in the fact that in addition to these, the answers given by the latter during the interrogation are also compared.

Indeed, the most recent developments in the jurisprudence of the American courts⁴⁹⁴ emphasise precisely the 'individualising' moment, precisely in order to avoid and overcome the possible (and risky) violation of the right to an individualised sentence. The software's operating process results in the risk score that is calculated by the machine: this is nothing but the result of a mixed-type assessment. It is so defined because it takes into account not only general statistics, but also the criminal profile of the defendant, which is reconstructed on the basis of the results of the questionnaire submitted to it.

Therefore, the response of the algorithm has a function that can only be defined as orientative of the magistrate's choices; in fact, the judging body is left with the discretion to decide whether or not to adhere to the index of dangerousness that is calculated; in the event that it decides to agree with it, it can assess the appropriateness of the same in determining the greater or lesser penalty.

In these terms, a system structured in this way, which therefore allows the judge to draw from the results of a mixed algorithmic assessment, adopting in advance all the necessary precautions⁴⁹⁵ to ensure respect for the defendant's defence guarantees, would be worthy of

⁴⁹³ Thus, in a critical key, M. REDMAYNE, *Character in Criminal Trial*, Oxford, 2015, 258.

⁴⁹⁴ Non a caso si prende in considerazione il caso *E. Loomis v. Wisconsin Supreme Court*, 881 N. W.ed 749 (Wis.2016), cert. Denied, 137 S. Ct. 2290, 2017.

⁴⁹⁵ On this point and the protections, see Chapter 4.

consideration even for the Italian criminal justice system⁴⁹⁶. Undoubtedly, as will be seen in greater detail in the next chapter, there will probably have to be a rethinking of the logic and structure of the prohibition of criminological expertise laid down in Article 220 of the Code of Criminal Procedure.

This is explained by the fact that in the current Italian legal framework, this very provision represents the obstacle and perhaps the limitation to the introduction of predictive algorithms in the Italian legal system.

Invero, il limite contenuto al secondo comma dell'art. 220 c.p. esprime la diffidenza del legislatore nei confronti del modello che si basa sull'evidence based sentencing⁴⁹⁷.

Indeed, the limitation contained in the second paragraph of Article 220 of the criminal code expresses the legislator's distrust of the model based on evidence-based sentencing.

In this respect, it is worth reflecting on two aspects: firstly, the decision to entrust a judge with the accurate weighing of all the symptomatic indices of the offender's capacity to commit offences (as provided for in Article 133(2), e.g. character, individual and social living conditions) is extremely difficult.

What comes to the fore and is the critical point on which the reasoning on the need to 'improve' such a choice is based is that the judicial assessment of these elements is resolved, in most cases, in a first 'impression' that the judge has; the same may be positive or negative on the impact that the judging body has on the personality of the defendant, which is undoubtedly conditioned by a series of elements and factors more or less explicit on the personality of the offender (first of all may concern the human conditionings that the judge possesses as a human being that pertain to the first impact and the evaluations that every human being makes).

Indeed, the idea that one decides to delegate in toto to a judge the need to know in full the character and personality of the offender would require the availability of means of investigation that would in themselves be prohibited in any case and in the absence of which he would find himself deciding with elements in his hands of c. so called 'ready-made solution' (such as, for example, previous convictions that have become final, the various factual circumstances that have emerged in the investigation or other elements from which he thinks he can derive the subject's capacity to commit offences, following an inductive-presumptive type of reasoning.

⁴⁹⁶ An attempt will be made in Chapter 4 to explain the reasons and grounds for optimistic consideration of the possible application and intrusion of these instruments into the Italian legal system.

⁴⁹⁷ This model clashes with the model of pure discretion. An attempt will be made to explore this in more detail in the following Chapter.

Such a method, as will be considered later, undoubtedly lacks objectivity and, one might even say, completeness of judgement, also because the judging body, in most cases, finds itself (on the basis of the available data) having to select only those elements to which it believes it must give a higher or lower value.

On the contrary, the possible use of algorithmic-actuarial risk assessment tools and techniques of the 'mixed' type could help to make the analysis entrusted to the judge more objective, offering him a valid guide to be able to direct the exercise of discretionary power⁴⁹⁸.

In these terms, precisely the presented model of evidence-based sentencing could be preferred or supported as an introduction also for another reason: for the trust it places in the 'more scientific' assessment of the offender's personality⁴⁹⁹.

In conclusion to this reflection, it should undoubtedly be noted that the progress that has been made by the psychological sciences in recent years should lead to a rethinking of the closed attitude hitherto displayed. On the other hand, in the criminal trial, science still has a marginal role and is not taken into consideration as an incontrovertible fact, but as a set of rational, albeit probabilistic⁵⁰⁰, knowledge. However, in the margins of the considerations made, while not questioning the purely probabilistic value of an investigation into the personality of the defendant, the results of the assessment would still be more objective than those derived from presumptive reasoning. These considerations lead one to reflect on the fact that one cannot fail to recognise that scientific and technological progress has rendered the model of discretion partly 'obsolete' even in the commensuration of punishment, which yields and should yield in the face of a model of conditional discretion⁵⁰¹.

⁴⁹⁸ Not only would certain indices that tend to be neglected (such as the character and individual personality of the offender) be taken into account, but also those social factors whose possible conditioning effect (place of residence, family composition, etc.) is not known a priori. It is these elements that can be deduced from the statistics compiled at regional or national level, which form the basis of the functioning of the predictive algorithms in the US national systems.

⁴⁹⁹As is well known, criminal law has shown a progressive openness to psychological diagnostics, which nowadays is of central importance in the investigation of certain offences. We refer to the events characterising the crime of persecutory acts, so-called stalking (Article 612-bis of the Criminal Code) and torture (Article 613-bis of the Criminal Code).

⁵⁰⁰ Of this opinion, G. CANZIO, *Il dubbio e la legge* 18, who speaks of judicial ascertainment as the art of judging «reasoning under uncertainty», albeit «by probabilities».

⁵⁰¹ For the concrete implementation of such a model, the legislator should intervene in the provisions of the Code of Criminal Procedure by introducing an express reference to the possibility of using risk assessment tools and techniques, subject to the abolition of the limits on criminological expertise. In this regard, it would be sufficient to introduce in the opening of Article 220(2) of the Code of Criminal Procedure an express reference to the commensuration of punishment among the activities for which it is permitted.

9 Possible remedies: enhanced and explanatory justification of the new algorithmic indices

One of the major problems that has arisen and on which doctrine has focused is precisely the need to establish the boundaries within which an A.I. instrument in the Italian legal system can be applied.

The judge's subjection to the law thus assumes articulated contours that imply the need to devote greater attention to the motivation of judicial pronouncements, a privileged place for knowing and, therefore, ensuring transparency and falsifiability of the argumentative process followed by the judicial authority⁵⁰².

The most delicate issue to be addressed concerns precisely the effects of its application as a supporting tool for the judge, having regard to the principle of the judge's free conviction. What indeed appears to be most difficult to establish in the first place relates precisely to the difficulty in establishing and delineating rules for understanding what is the actual room for manoeuvre in the decision-making of a judge using an A.I. tool.

In particular, one wonders to what extent and whether the judge could deviate from the evidentiary evidence of the electronic computer and whether one must, instead, imagine a sort of bindingness that obliges the judge to follow the result in the decision-making phase⁵⁰³.

Therefore, it would be conceivable and conceivable plus the possibility that the judge would be free to weigh the known variables and to account for the particularities of the concrete case that possibly require departing from the result of the algorithm⁵⁰⁴.

Undoubtedly, criticism has already been raised by the doctrine concerning the possibility and risk that the algorithmic result ends up totally conditioning and limiting the judge in his decision; this is also explained by the fact that there is in general an enormous persuasion that technology has on human beings; in such a case, the risk would be that judges would make their decision depend on the predictive outcome provided by the algorithm, limiting themselves in their decision to validating its results. This phenomenon, in particular, is known in cognitive psychology as 'anchoring'⁵⁰⁵, which would derive from the tendency of humans to be called

⁵⁰² M. VOGLIOTTI, *La nuova legalità penale e il ruolo della giurisdizione. Spunti per un confronto*, in *Sistema penale*, 3, 2020, 60, emphasising the process of transition from legalistic legality to the 'new' legality.

⁵⁰³ In this case, one must bear in mind Article 22 GDPR, which, by prohibiting the adoption of purely automated judicial decisions unless authorised by consent, contract or the law of the Member States, would seem to exclude the binding nature of the evidentiary result offered by the software. On this topic, see Chapter V.

⁵⁰⁴ See G. TUZET, *L'algoritmo come pastore del giudice? Diritto, tecnologie, prova scientifica*, in *Medialaws*, 16th March, 2020.

⁵⁰⁵ In particular, it is a phenomenon whereby a human decision maker attaches a certain weight to a tangible and immediately available piece of data in a way that is potentially detrimental to the decision. Please refer on this point to S. ARCEIRI, *Bias cognitivi e decisione del giudice: un'indagine sperimentale* in *DPU*, 2, 2019. For

upon to make decisions under conditions of uncertainty and to rely on the available evidence, regardless of its scientific validity⁵⁰⁶. In particular, the tendency of humans to unconsciously and irrationally place unconditional trust in technologies, which are deemed objective and trustworthy 'merely because ... they are technologies' is referred to as the 'automation fallacy'⁵⁰⁷.

Obviously, this issue is not extraneous to criminal law⁵⁰⁸, insofar as it relates to the age-old debate concerning the connection between scientific evidence and the judge's decision, which arises whenever, pursuant to Article 220(1) of the Code of Criminal Procedure, expert evidence is admitted and the judge relies on the expertise and professionalism of third parties, in the capacity of experts or technical consultants.

Indeed, a possible remedy has been thought of to prevent the judge from being subjected to 'external' conditioning by expert knowledge and the fascination of technology and artificial intelligence tools, as would be the case, for example, in the hypothesis of recourse to predictive software, in which it would be possible to find a natural guarantee 'antibody' in terms of the motivation of the judgment. In particular, it is considered that the judge should explain in the grounds the evidence adduced in support of the decision, without however 'flattening' his judgment on the IT findings, arguing the specific reasons that led him to take them into consideration and, possibly, to consider them preferable to other elements and evidence available⁵⁰⁹.

As has already been hypothesised, in such a case, one should assume the idea that the judge's reasoning should focus on the following aspects: reliability (i.e. verifying that using the same algorithm and method several times leads to the same result); validity (i.e. verifying that the result obtained reflects the state of affairs); generalisability (i.e. verifying that the result obtained is also applicable to other similar cases) and, finally, credibility (i.e. proving that the procedure and results obtained cannot be falsified). In this way, the result of the predictive software would be 'filtered by the human mind, which will have to interpret the data that the

example, when it is necessary to make a numerical estimate (e.g. the market value of a house), people tend to rely on the first piece of data available (e.g. the list price). The final estimate tends to 'anchor' to that initial value.

⁵⁰⁶ Indeed, predictive software is 'a convenient shelter for the judge who, hiding behind the score, could fail to consider all the particularities of the case and, as an immediate consequence, fail to give adequate reasons for the commensuration of the penalty, thus J. GERARDS, *The fundamental rights challenges of algorithms in Netherlands Quarterly of Human Rights*, 2019, 37(3):205-209.

⁵⁰⁷ P. CAMOGLIO, *Prefazione*, in J. Nieva-Fenoll (ed), *Intelligenza artificiale e processo*, trad. it. Turin, X-XVI.

⁵⁰⁸ Thus, on this point the critical considerations of A. MANNA, *I rapporti tra sapere scientifico e sapere giudiziario*, in *Cass. Pen.*, 2009, 3633, who agrees with the stigmatisation of the judge as a 'bureaucrat flattened on the knowledge of the expert' who must constitute, on the contrary, 'a valid, qualified support for the judge in the evaluation and decryption of the elements of (scientific) evidence, but must not become the verdict on the evidence, otherwise the feared risk of a technicalistic drift would materialise, eclipsing in toto the stated principle of free conviction'.

⁵⁰⁹ On the different burden and obligation to state reasons, see Chapter 5.

survey provides⁵¹⁰, being externally controllable through the logicity and coherence of the reported arguments.

In other words, therefore, the judge would decide after having 'heard' the algorithm.

In this way, therefore, while not knowing the limits of applicability, predictive software could find its way into the criminal trial process, especially in those areas that not only require forward-looking analysis and evaluation but are also particularly exposed to cognitive distortions.

In conclusion, there are several critical issues and the most delicate points on which it is necessary to reflect; on closer inspection, the algorithmic assessment of dangerousness undoubtedly leaves many questions open, touching on several aspects: from the defendant's guarantees, to the reviewability of the final result, on the falsifiability and scientific fallibility of the software, on the difficult reliability of the inputs and outputs, on the residual duty of motivation entrusted to the judge, on the discriminatory effects due to empirical generalisations and to the social and economic conditioning facts that are processed by the algorithm⁵¹¹.

Thus, ensuring due guarantees for the defendant, such a resource in this case could prove to be an invaluable opportunity to be seized.

⁵¹⁰ See a G. F. RICCI, *Nuovi rilievi sul problema della "specificità" della prova giuridica*, in *Riv. Trim. dir. Proc. Civ.* 2000, 1129.

⁵¹¹ Some of these aspects were addressed by the Wisconsin Supreme Court in the famous Loomis case, with a ruling that, while confirming the legitimacy of the use of these instruments, issued a series of warnings to the judges on the merits on the caution to be taken in their use. Thus, on the point Wisconsin Supreme Court, *State v. Loomis*, case 2015AP157-CR, Judgement July 13th, 2016, in *Harvard Law Review*, 2017, vol. 130, 1530 ss.

Chapter Four

A second field of application: the choice of the best penalty treatment Sentencing which transforms. Between risks and benefits

SUMMARY: 1. Methodological premise: between the judge's decision and risk assessment tools. – 1.1. The continuation of a premise: between the decision and the future with risk assessment tools. – 2. The most delicate phase left to the judge: the choice on the commensuration of the penalty. – 2.1. The criteria and the different types of penalty: how much the judge's discretion is gradually affected. – 2.1.1. Initial reflections on the Italian discretionary system. – 3. Focus: the application of the sentence and the judge's discretionary power. – 4. American judicial practice and the use of risk assessment: a special *focus* on the selective incapacitation movement theory and evidence-based sentencing. – 4.1. *Segue*: the penalty phase in the US system. – 4.2. How actuarial risk assessment came about. – 4.3. *Compass*: the Loomis case and the Wisconsin Supreme Court decision. – 4.4. The peculiarity of the decision: the decisional 'double phase' in the choice of penalty treatment. – 4.5. An application overview of the United States: the case of Virginia. – 5. From risk assessment in recidivism to sentence commensuration: why algorithms fit into sentencing. – 5.1. The intersection of two provisions at the stage of assessing the penalty treatment. – 6. Criminal discretion in the Italian legal system: the difficult framing. – 6.1. Sentence commensuration and criticised discretion. – 6.2. The 'capacity to commit offences' as an assessment that forces one to look into the future. – 6.3. The problem of prognostic evaluations. – 6.4. The answer to a question: why *prognosis* is considered so important in the choice of sanction treatment. – 7. Limits and differences in algorithmic evaluation in the Italian penal system. – 7.1. The limits posed by Article 220 of the Code of Criminal Procedure: is the principle in crisis? – 8. Discretion and its combination of constrained and controlled in the criminal justice system. – 8.1. A new constrained discretion: the judge's free conviction in the face of new probative evidence. The Weakness of Articles 132 and 133 of the Criminal Code. The Praxeological Guidelines on Discretion. – 8.1.1. The problem of the weakness 'in the dark' of prognostic judgements. – 8.2. The concept which returns: social dangerousness within Article 133 of the criminal code. – 8.3. Problems and the first emerging evidence on the phenomenological level. – 9. The human decision and the technological decision: a surmountable opacity? An adversarial 'technicalisation'. – 9.1. The algorithm in the decision-making phase: what benefits and towards what future? – 10. The paradigmatic value of Article 133 of the Criminal Code: inadequate *criteria*? – 10.1. The *ethicality* of human judgement and its ineradicable subjective components. – 11. Concluding remarks: drawing conclusions on risk assessment.

1 Methodological premise: between the judge's decision and risk assessment tools

As already mentioned, in the most recent studies on the applications and use of A.I. within six criminal justice systems, the possible use of predictive algorithms in the service of the judge⁵¹² to calculate (in addition to the dangerousness of an individual) the risk of re-offending,

⁵¹² In fact, the idea of the automaton-judge was born with the formal conception of law developed by Montesquieu, who imagined the organ devoted to the application of laws as an 'inanimate being', 'depersonalised', with the task of being a mere executor of the provisions drafted by the legislative assembly elected by the people. According to the mechanistic model conceived by the philosopher and affirmed by the French Revolution, the decision should have consisted in a logical operation of subsumption of the concrete case to the abstract one envisaged by the norm, in an 'automatic' application of simple deductive syllogisms. Legal logic would therefore have coincided with formal logic and the judge, without deploying any discretion and any power, and distancing himself from his own opinions, his own ethical convictions and his own emotions, would have behaved just like a machine. As

which in itself is also related to the first type of assessment, both in the application of precautionary and alternative measures, and in the commensuration of punishment, is increasingly emerging⁵¹³.

Although there has been considerable focus on using risk assessment algorithms in rehabilitation and especially in pretrial decision-making, they have recently drawn attention for their use in sentencing⁵¹⁴.

A first question that needs to be answered, and from which all the reflection moves, concerns the peculiarity of 'decision', i.e. the macro-interrogative on: what does it mean to decide?⁵¹⁵

Indeed, in the cognitive sciences⁵¹⁶, the concept of judgement is distinct from that of decision-making. Judgment is more understood as 'a cognitive experience that leads to the formation of a conviction'. Conversely, a 'decision-making' process is a cognitive performance that leads to making a choice. Clearly, the matter changes when one approaches the judicial decision, which in fact encompasses both of the above-mentioned aspects: on the one hand, it implies a judgement since it assesses the existence of a quality or condition (in the case, for instance, of the decision on guilt/non-guilt) of the defendant; on the other hand, it triggers a real decision-making process, through which the possible consequences are established (the *quantum* of sanctions, the benefits provided for by law, the possible measures to be imposed)⁵¹⁷.

Norberto Bobbio wrote, however, 'a lot of water has passed under the bridge since the era of so-called legislative fetishism, and no one seriously believes in the judge as an automaton anymore': it is now universally accepted that the activity of jurisprudence is always inevitably creative, and that the mechanistic ideal of the judge *bouchede la loi* is unrealisable. See, F. C. GASTALDO, *Il giudice-robot: l'intelligenza artificiale nei sistemi giudiziari tra aspettative ed equivoci*, March 2021. See also on the inevitably creative role of case law in the application of the law, G. TARELLO, *L'interpretazione della legge*, in AA.VV., *Trattato di diritto civile e commerciale*, Milan, 1980, 1 – 99. The subject is, moreover, a complex one: on the activity of the judge understood, otherwise, as the 'invention' (inventio) of law derived from the regulatory tradition in force in a society, P. GROSSI, *L'invenzione del diritto*, Bari-Roma, 2017; N. BOBBIO, *Giusnaturalismo e positivismo giuridico*, Bari-Roma, 2011.

⁵¹³ A. M. MAUGERI, *L'uso di algoritmi predittivi per accertare la pericolosità sociale: una sfida tra evidence-based practices e tutela dei diritti fondamentali*, in *Arch. Pen.*, No. 1, 2021.

⁵¹⁴ See, J. ANGWIN, *Make algorithms accountable*, *New York Times*, 1st August 2016.

⁵¹⁵ Così, S. QUATTROCCOLO, *Per un'intelligenza utile al processo penale*, 390.

⁵¹⁶ For example, cognitive psychology and behavioural economics are the branches that have most thoroughly studied the subject; thus, on the point G. CEVOLANI-V. CRUPI, *Come ragionano i giudici: razionalità, euristiche e illusioni cognitive*, in *Criminalia*, 2017, 181 ss.

⁵¹⁷ In fact, this pattern as described is much more clearly evident in legal systems that provide for a sharp caesura between fact finding and sentencing. This distinction has historical roots and has often been associated with the presence of a jury trial, a feature certainly more common in common law systems. For a more specific historical and comparative analysis, it is recalled that France introduced the jury institution in criminal trials in the late 18th century, due to English influence. Although, at that time, the philosophical debate dissipated throughout Europe, the jury model has never been predominant on the continent. In addition, the 20th century rise of illiberal regimes, both fascist and Bolshevik, led to the replacement of surviving juries with professional or mixed courts in many continental countries: on this point, see A. PADOA SCHIOPPA, *La giuria penale in Francia*, Milan, 1994, 7 ss.; J. H. LANGBEIN, *The English Criminal Trial Jury on the Eve of the French Revolution*, in A. Padoa Schioppa (eds), *The Trial Jury in England, France and Germany, 1700-1900*, Berlin, 1987, 16 ss.; S. THAMAN, *Should criminal juries give reasoning for their verdicts? The Spanish experience and the implications of the European Court of*

In the light of the US experience and the exposition of the Artificial Intelligence tool models taken into consideration, it is necessary, at this point of the discussion, to evaluate the limits and possibilities underlying the possible introduction, in the Italian justice system, of predictive software capable of calculating and supporting the judge in the extremely delicate sentencing phase, in the choice of the *quantum* of sanctions. The issue is as complex as ever if one considers that, firstly, the most suitable tool to respond to such purposes must be identified and, secondly, within what boundaries with a view to assessing what actually can be considered the advantages deriving from such application.

It is undoubtedly necessary to proceed step by step and assess the practical-operational 'feasibility' of such a solution, having regard to the scientific knowledge currently available and whether computer experts and technicians can actually develop such a predictive programme; if so, to verify and identify what the inputs to be entered and the programming technique, if any, might be. With regard to the latter, the issues arising from the so-called 'algorithmic question' and the availability of the necessary data must be assessed and resolved.

Once the field has been cleared of questions of a technical-informatic nature, the analysis will shift to a properly legal level, in terms of the impact that the adoption of such a model of predictive justice might entail in our legal system. Precisely by following the line of this perspective, one can see how advantages and risks, virtuous effects and limits, guarantees and unknowns overlap. It will then be necessary to assess and verify the sustainability and usefulness that such an IT support tool might be able to ensure and as a (possible) corrective to the cognitive distortions that are recurrent in the choice of penalty treatment.

In the first place, it would be necessary to understand whether it is technically possible to design a software that would allow, on the one hand, to 'calculate the foreseeability of a given offence event (and thus the possibility of recidivism) and, on the other hand, on the wave of the first assessment, a software that would be able to assist the judge in assessing all the elements available on the offender for the purpose of identifying (with a match) what could be the best sanctioning treatment to be imposed on the offender in order to fully realise the re-educative purpose of the punishment.

Human Rights decision in Taxquet v. Belgium, in A. Petrova (ed), Festschrift für August Nacke, 2016, 338-385. Recently, the European Court of Human Rights examined the solutions adopted by Council of Europe countries regarding juries in criminal matters. In *Taxquet v. Belgium*, the Court assessed that there are three different approaches to the issue. Countries that do not have (and some have never had) a jury. A large group of countries that provide for a mixed jury, where 'the professional judges and the jurors collectively determine all questions of law and fact, the issue of guilt and the sentence'; finally, a number of systems that are based on the institution of the jury as the judge of fact, which mostly delivers unreasonable verdicts.

From this point of view, it is therefore essential to identify what should be the moment of programming. Following this direction, it is considered that (unlike the cases in which it is all the more necessary to programme the software as a preventive measure) the design of the software could also be commissioned later when the sanctioning treatment is chosen. This is where the necessary data concerning the individual should be input (once the process of recognising the individual's criminal liability has already been completed).

Lastly, it would remain to be considered (it will not be possible to fully analyse this question here) the possible design costs; in this direction, it is believed that in order to cope with these burdens and the high design costs, one could imagine preparing a software-model, which is given back to the judicial authority, not developed for the forecast calculation of an event or a specific risky activity, but adaptable to the concrete case from time to time. However, it would remain a software that is prepared and constructed ad hoc to answer specific questions.

1.1 The continuation of a premise: between the decision and the future with risk assessment tools

In order to make a brief systematic overview, it should be noted immediately that Title V of the Italian Criminal Code deals with the "modification, application and enforcement of sentences" and includes provisions concerning, on the one hand, the discretionary power of the judge in the commensuration of the penalty, the legal criteria for exercising this power, the calculation of the principal and accessory penalties (Chapter I) and, on the other hand, on the subject of enforcement, the remuneration of the work of the convicted and the postponement of the expiry of the sentence⁵¹⁸. Indeed, a central role is undoubtedly played by both the determination of the penalty at the judicial stage⁵¹⁹, i.e. the complex and delicate assessment operation of commensuration of the sentence, at the disposal of the judge of cognition at the time of sentencing, in terms that are adequate, consistent and proportionate to the particularity and individuality of the concrete case, which marks the transition from the abstract level of the sentence to the concrete level of the penalty imposed on the offender⁵²⁰.

⁵¹⁸ The precariousness of the systematics and title of the title, in particular of its two chapters, has been strongly criticised, on the one hand because of the so-called "remoteness of the location" between the individual provisions relating to individual penalties and those of Chapter I, which are intended to regulate their judicial enforcement, and on the other hand because of the anachronistic placement of the few provisions of Chapter II, which have survived the legislative upheavals and concern a subject, the enforcement of penalties, which, in view of the reforms that have taken place in the meantime, should be more appropriately reserved to a specialised discipline: thus on the point M. ROMANO-GRASSO, *Commentario sistematico del codice penale*, Vol. 2, Milan, 2012, 314.

⁵¹⁹ See, T. PADOVANI, *La pericolosità sociale sotto il profilo giuridico*, 319.

⁵²⁰ On this point, F. PALAZZO, *Corso di diritto penale. Parte generale*, Turin, 2018, 580.

As anticipated, 'the area in which judges have most traditionally been called upon to assess the defendant's future conduct is that of sentencing'⁵²¹. In this regard, however, it should be borne in mind that many continental systems, unlike the common law tradition, do not provide for a clear distinction between affirmation of guilt and quantification of punishment, the former, unmotivated, referring to the ascertainment of the facts and the fulfilment of the burden of proof, the latter, motivated, aimed at quantifying the penalty⁵²². Although the philosophical debate spread, with the wind of revolution, throughout Europe, the popular jury never became the predominant model on the continent. Moreover, 'the emergence in the 20th century of illiberal regimes, both fascist and Bolshevik, led in many continental countries to the replacement of popular juries - where they existed - by professional or mixed bodies'⁵²³. Recently, the European Court of Human Rights examined the solutions adopted by the various member states of the Council of Europe with regard to juries in criminal matters⁵²⁴.

We will therefore try, in the following paragraphs, to start from a comparative analysis⁵²⁵ in an attempt to show the current situation in the USA in order for you to assess, starting from the intrinsic characteristics of the Italian decision-making phase, how one can imagine incorporating such I.A. instruments.

To this end, precisely in the concluding part of this chapter, the possibility of proposing a new model, starting from the current one of judicial discretion, to a constrained discretion

⁵²¹ S. QUATTROCCOLO, *Sui rapporti tra pena, prevenzione*, 271.

⁵²² This distinction has historical roots and has often been associated with the presence of a jury trial, and it is certainly a feature shared by most of the systems twinned on English common law. At the time of the French Revolution, France too moved towards a system of trial by jury, precisely as a result of the English influence, for which Montesquieu had paved the way in his monumental work *De l'Esprit des Lois*, through his praise of the 'judgment of peers', emblematic, according to the philosopher, of procedural fairness.

⁵²³ S. QUATTROCCOLO, *Sui rapporti tra pena, prevenzione*, 271.

⁵²⁴ In the well-known *Taxquet v. Belgium* case, the Court recorded three different approaches to the issue. Alongside countries that do not have (nor have they ever had) a jury, a large group³⁸ provides for mixed juries, in which "the statutory judges and the jurors collectively determine all questions of law and fact, the issue of guilt and the determination of punishment", while a third group of countries opted for a 'pure' popular jury system³⁹. Although the differences within this third group are many, "the general rule seems to be that verdicts reached by a traditional jury are not motivated. This is the case in all the countries concerned, with the exception of Spain and Switzerland (Canton of Geneva)". Si rimanda al caso CEDU, Gr. Ch., 16.11.2010, *Taxquet c. Belgio*, in www.echr.coe.int.

⁵²⁵ It should be noted at the outset that, as regards the 'caesura' between guilt and punishment, Belgium is an interesting continental example, providing for two separate moments of deliberation⁴⁰. In other legal systems, belonging to both the first and the second group, (e.g. Germany, Spain, Italy, Switzerland, Portugal, among others), the court of first instance, while following a logical concatenation that obviously puts the decision on guilt before that on the quantification of the penalty, pronounces without a break - and with an obligation to state reasons - on guilt and sentence. Interestingly, the caesura between these two decision-making moments has been theorised as a crucial instrument of modern criminology, especially by Marc Ancel and the so-called New Social Defence group. And France itself, for example, has recently reinforced the distinction - functional, rather than temporal - between conviction and punishment, with the law of 15 August 2015 on the individualisation of punishment and the subsequent law of 23 March 2019, implementing the former.

where, however, a significant role is assumed and attributed to predictive algorithms and, more generally, to actuarial assessment methods of the offender's capacity to commit offences.

What certainly has to be considered, in the light of what has been said in the previous chapter, concerns the fact that the analysis of dangerousness and the abandonment of penal subjectivism that looked at the perpetrator and moved away from the fact, is a conception that has remained anchored and typical of totalitarian regimes.

In conclusion, to this premise, it seems useful to note how, with regard to the relationship between offence and author, the history of criminal law has always, over the years, oscillated between a: criminal law of the pure fact, a criminal law of the author, and a mixed criminal law of the fact and the author's personality.

It would seem almost as if the debate centred on the introduction of such instruments at certain precise stages of the criminal process in order to obtain as complete an analysis as possible of the characteristics of the subject, could move closer to the last vision which, while remaining anchored to the guaranteeing principle of the fact as the inescapable basis of all criminal consequences, at the same time takes into account the inescapable need to assess the personality of the offender, if only in order to determine the type, quantity and duration of the applicable criminal consequences.

However, to date, it is considered appropriate to point out that in criminal systems of protectionism, the type of offender is taken into account and how it can be a criterion for assessing the offender's capacity to commit offences in order to be able to graduate the punishment in the best possible way, since the meeting point between the seriousness of the offence and the offender's personality must be found⁵²⁶.

Although the focus in this paper will be more on the provision set out in the second paragraph of Article 133 of the Criminal Code, there are several provisions scattered throughout the code - which relate more to the enforcement phase of the sentence - that are themselves based on prognostic judgments on dangerousness.

Indeed, these provisions are undoubtedly united by a common feature: the prognostic judgment and the relevance of the assessment of recidivism. This represents the real interchange of the penalty system that then allows the judge to modify the response of the offence. In fact, if he considers that the offender (or the defendant) does not then commit other offences, the penalty threatened (in the abstract) and identified (in concrete) will then undergo a transformation in content.

⁵²⁶ F. MANTOVANI, *Manuale di Diritto penale*, Milan, 2020, 604 ss.

In fact, one cannot fail to note the importance and centrality of prognosis, which represents the fundamental passage through which the punitive (or para-punitive) option capable of best achieving the purposes of positive special prevention is identified. Indeed, it is a judgement that is not only indispensable in a teleological perspective (in fact, special prevention cannot be realised if not with an eye to the future) but which, on closer inspection, is constitutionally obligatory since only through the estimation of the possibilities of re-education of the convicted person can the constitutional principle dictated by Article 27, paragraph 3, of the Italian Constitution be complied with.

Recently, risk assessment tools have also assumed a significant role in the delicate phase of the trial. In fact, judges increasingly rely on the outcome of the algorithm to guide their decisions on determining the sentence to be applied⁵²⁷.

2 *The most delicate phase left to the judge: the choice on the commensuration of the penalty*

On closer inspection, one of the most delicate phases after the decision on a person's guilt relates to the judge's assessment of the quantum of the penalty. This refers to the moment when the judge has to perform an operation consisting: at first in the identification of the type of penalty to be applied for the offence committed and, at a later stage, in the determination of the quantum of punishment⁵²⁸ to be imposed concretely on the offender. This assessment unfolds and is carried out on the basis of a sentence range that makes the judge move between a minimum and a maximum sentence (predetermined *ex lege* by the legislature).

Parallel to this premise, moreover, as already mentioned, the criminal sanctioning system is pervaded today by institutes whose application requires, more or less expressly, the formulation of a criminological prognosis⁵²⁹ by the judge, i.e. a discretionarily bound and fatally probabilistic assessment of the future conduct of the accused or convicted person.

⁵²⁷ In some jurisdictions, the use of the result provided by such tools is strongly encouraged by law. Such as in Hampshire, Pennsylvania, Arkansas and Vermont. For example, in Oklahoma the use of "assessment and evaluation instruments designed to predict risk of recidivism to determine eligibility for any community punishment" is mandated.

⁵²⁸ On the notion of commensuration of punishment see in particular, G. BELLAVISTA, *Il potere discrezionale nell'applicazione della pena*, 1939, in *Il Tommaso Natale*, 1975; G. BETTIOL, *Pena retributiva e poteri discrezionali del giudice*, in *Riv. it. dir. pen.*, 1941, 109 ss.; E. DOLCINI, *La commisurazione della pena. La pena detentiva*, Padua, 1968, 4; ID, *La commisurazione della pena: spunti per una riforma*, in *Riv. it. dir. proc. pen.*, 1981, 34 ss.; Id., *Potere discrezionale del giudice* (dir. proc. pen), in *Enc. dir.*, XXXIV, Milan, 1985, 745 ss.; G. FIANDACA-E. MUSCO, *Diritto penale. Parte generale*, 5th ed., Bologna, 2007, 703 ss; S. LARIZZA, *La commisurazione della pena. Rassegna di dottrina e giurisprudenza*, in *Riv. it. dir. proc. pen.*, 1982, 596 ss.; V. MILITELLO, *Prevenzione generale e commisurazione della pena*, Milan, 1982; A. PAGLIARO, *Commisurazione della pena e prevenzione generale*, in *Riv. it. dir. proc. pen.*, 1981, 25 ss.

⁵²⁹ On this point, G. KAISER, *Criminologia*, Milan, 1985, 137.

The debate today is certainly, as already anticipated, drawing attention to how numerous and varied are, in reality, the hypotheses in which the criminal judge is called upon to make a predictive judgement of the defendant's behaviour, which is then decisive for the latter's personal freedom, outside a clear framework of elements to be used later for the decision⁵³⁰.

On closer inspection, in the study of this particular and delicate phase entrusted to the judge, it was realised that despite the considerable theoretical and practical importance of prognostic judgments in a punitive model that pursues the purpose of prevention, the subject of prognosis in the penalty system does not appear to be much investigated⁵³¹. This premise is strengthened even more on the subject of commensuration of punishment; on the contrary, reflections on security measures have always turned their attention to prognosis, which has certainly been the subject of a wider and more in-depth treatment: on the one hand, not only because the ascertainment of social dangerousness constitutes one of the prerequisites for the application of security measures, but even more so because the prevention of recidivism, through care and control, represents the very essence of the security track of the Italian penal system.

It is precisely the delicate phase of the choice of sanctioning treatment that is one of the areas in which judges are traditionally most often called upon to assess the future conduct of the defendant. Indeed, the essence in itself of this type of decision implies a twofold type of assessment: one addressed to the facts that have occurred and to the type of sanctioning response that the legal system decides to impose on a given subject (disvalue of the fact and of the event); on the other hand, the choice of treatment cannot but address the future in a twofold sense. In fact, in the first place, one must look at the risk that the subject may commit a crime (of the same nature and species) or even a different crime (and in this case it is strictly connected to the dangerousness of the individual), but at the same time, one must look, with a view to the full realisation of the principle expressed in Article 27(3) of the Italian Constitution, at what the best treatment for the subject may be, with a view to a punitive and at the same time re-educational response.

In fact, it is no coincidence that prognoses operate both in the choice and commensuration of the sentence, and during its execution and throughout its duration⁵³².

⁵³⁰ S. QUATTROCCOLO, *Sui rapporti tra pena, prevenzione*, 272.

⁵³¹ See L. MONACO, *Prospettive dell'idea dello 'scopo' nella teoria della pena*, 150.

⁵³² On closer inspection, prognostic judgements are of decisive importance in relation to numerous alternative responses to custodial sentences in the execution phase. One thinks, in this regard, of alternative measures to detention. The sources and tools that offer the judge, also through the work of the prison staff, a potentially complete cognitive picture of the offender and of the predictive factors of recidivism in the case in question are to be examined in depth. Here too, as we shall see, the prognostic assessment of the risk of reoffending is carried out on the basis of maxims of experience and the personal intuition of those called upon to decide, without sufficient recourse to empirical and scientific knowledge. This perspective would undoubtedly deserve an

It therefore appears necessary to reflect on this aspect of the decision and why it is closely linked not only to the historical fact but also to the future, a decision that therefore looks forward⁵³³.

In this regard, over and above the differences on the separation or union of the moment of sentencing and that of quantifying the penalty, it is worth noting how this last aspect represents the moment in which all the competing penal ideologies come into confrontation: the lack of consensus on the very function of the penalty, in sentencing, in fact emerges very clearly. It is no coincidence that the key terms on which the debate on sentencing and sentencing commensuration (also read with a view to making room for risk assessment tools) then mostly focuses are precisely: deterrence, punishment, rehabilitation and dangerousness. These represent the main competing ideologies that collide and intersect with each other⁵³⁴.

Indeed, although a trial aims to reconstruct an event that occurred in the past - in order to establish the possible guilt of the defendant - penal systems also tend to attach importance, in the judgment, to his possible future behaviour⁵³⁵.

In other words, the very debate and reflections on the weight that the future conduct of the defendant can or should have in the definition of the pending proceedings, represents a classic of modern specialist literature (today it is a theme that has, so to speak, been revived even within criminal law scholars).

Precisely for the reasons set out, it is considered necessary to focus the debate on the sentencing line that we have decided to retrace in this paper, since it is considered possible to argue, albeit in general terms, that the relationship between the risk of the future commission of a crime and the personal liberty of the defendant is deeply rooted in the design of contemporary criminal justice (or in western legal culture) in which the general-preventive function is inherent in the penal system⁵³⁶.

Not only that, but every time the space recognised in a legal system for special prevention grows⁵³⁷, predicting, measuring, indicating the possibility and probability of criminal behaviour, more specifically actually violent behaviour, becomes crucial even for the purposes of determining punishment, which is why today we also speak of predictive sentencing.

independent analysis and investigation of the possible application implications of such instruments in the penalty enforcement phase.

⁵³³ On this point, S. QUATTROCOLO, *Sui rapporti tra pena, prevenzione*, 16 ss.

⁵³⁴ A. NORRIE, *Punishment, Responsibility, and Justice: A Relative Critique*, Oxford, 2014, 335

⁵³⁵ See, S. QUATTROCOLO, *Sui rapporti tra pena, prevenzione*, 272.

⁵³⁶ Indeed, on the role of general-prevention in the determination of punishment remains fundamental E. DOLCINI, *La commisurazione della pena*, Padua, 1979, 153 ss.

⁵³⁷ E. DOLCINI (et oths), *Il diritto alla Speranza. L'ergastolo nel diritto penale costituzionale*, Turin, 2019, 5.

On closer inspection, it is precisely for these reasons that risk assessment (born out of psycho-criminological theories) is one of the areas of criminal justice in which the digital revolution has been most exploited and utilised.

Indeed, risk assessment tools are also based on the statistical analysis of data sets, collected from relevant population samples, in order to build models that assess the personal level of potential offence or recidivism of individuals⁵³⁸. As already mentioned in the previous pages, precisely such statistical and algorithmic tools have enabled researchers over the years to empirically identify risk factors significantly associated with the commission of criminal and violent behaviour.

In this scientific field of study, we have witnessed, especially in recent years, the proliferation of focus on them as computing power has increased and the processing of vast amounts of data has expanded. Indeed, parallel to this unstoppable evolution has been the evolution of predictive risk models into computational predictive risk models, with a clear expansion of reference databases, speed of analysis and consistency of results⁵³⁹. Therefore, even in those jurisdictions where predictive sentencing has never been the rule, it has become necessary to question the nature of risk assessment in order to be able to understand what the limits and benefits might be in order to be able to propose their introduction in criminal proceedings.

In the first place, it must be borne in mind that the criminal section in particular performs its function by itself in three stages: the threat, which is proper to the law, the infliction, which falls within the judge's choice and activity, and the third, conclusive stage, which relates to the moment of execution. In the second step, the operation of commensuration of the penalty comes into play, which is then closely linked to the choice of sanction that at this point begins to acquire concreteness⁵⁴⁰.

The question of the commensuration of punishment is one of the thorniest and among the most peculiar, from which one can deduce the continuous search for a balance between legality and social defence, between certainty and substantial justice. A balance that is as difficult as ever to find since it is undoubtedly connected to the very essence of the sanction that looks not

⁵³⁸ G. ZARA – D. FARRINGTON, *Criminal recidivism: explanation prediction and prevention*, 150.

⁵³⁹ On this point, A. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro della libertà*, 71 ss.

⁵⁴⁰ On the three phases in which the proper function of punishment takes place, G. BELLAVISTA, *Il potere discrezionale nell'applicazione della pena*, 141.

only to the 'reprimand' for the deed committed but also to the future in its essential purpose of re-educating the convicted person⁵⁴¹.

In conclusion, as we know, the criminal justice scenario imposes, at all times, a necessary synthesis between different objectives and values, such as social security, human dignity and re-education. However, the balance between these values is not always fixed but rather varies over time, in parallel with the evolution of society and political and social sentiments.

Therefore, in this context, the boundary between repression and prevention becomes blurred: despite, in fact, the presence of a double sanctioning track between punishment and security measures, nevertheless, the social dangerousness of the individual (which basically remained countered and confined by the application of security measures) has ended up today assuming relevance also for sanctioning purposes, becoming an index for the commensuration of punishment.

As already mentioned, this type of assessment per se requires a prognostic type of analysis referred to the judge. Clearly, this type of assessment calls into play not only criminal law but also the empirical sciences⁵⁴²; despite this, the analysis in this paper certainly starts from the premise (and hence the necessary and indispensable reflection) that there is and has been remarkably little attention devoted to the role of prognostic judgments in the penalty system.

It is precisely this element that leaves open numerous problematic questions on the methods applied to make them; on the choice of factors to be taken into consideration, on their justification and verifiability and on the general principles governing them. For this reason, it was therefore decided to embark on this path in an attempt to assess the possible corrective measures and perhaps systems that could be introduced within the justice system to support the adjudicating body.

Thus, an attempt will be made here to turn our attention to the characteristics of the sentence commensuration phase, trying to identify the weaknesses and shortcomings of the assessment of the capacity to commit offences entrusted to the adjudicating body.

In other words, in this peculiar phase of the conclusive judgement, the judge has enough information about the fact and the offender; however, the prognostic judgement that is issued

⁵⁴¹ Still on the problem of the commensuration of punishment and the principles that underlie it, S. MESSINA, *La discrezionalità nel diritto penale*, Rome, 1947; G. VASSALLI, *Il potere discrezionale del giudice nella commisurazione della pena*, in *Primo corso di perfezionamento per uditori giudiziari*, II, Milan, 1958, 725 ss.; P. NUVOLONE, *Il ruolo del giudice nell'applicazione della pena*, in *Trent'anni di diritto e procedura penale*, II, Padua, 1969, 1558 ss.; T. DELOGU, *Potere discrezionale del giudice e certezza del diritto*, in *Riv. it. dir. proc. pen.*, 1976, 369 ss.

⁵⁴² As repeatedly stated by G. ZARA, *Valutare il rischio in ambito criminologico. Procedure e strumenti per l'assessment psicologico*, Bononia, 2016, 17.

is vitiated by a deficit of the knowledge necessary to formulate it, perhaps because, beyond the regulatory limitations encountered, there is also a constant cultural attitude of distrust or underestimation of the importance of prognostication in the penalty system.

2.1 *The criteria and the different types of penalty: how much the judge's discretion is gradually affected*

On closer inspection, after setting out the definition of the very delicate phase of the commensuration of punishment⁵⁴³, it is noted how, from the earliest doctrinal debates, the problem of the commensuration of punishment has been debated through different types of dogmatic solutions that differ from the alternating 'power of choice' between the judge and the law.

Indeed, we can see how this phase can see the prevalence of: absolute legality, a kind of 'free discretion' and constrained discretion.

With regard to the first type, one can see how legality and the law are in the foreground; indeed, these are systems in which the determination of the penalty, both in kind and in amount, is peremptorily and predeterminedly established by law. In this case, it is a sanctioning system in which the principle of legality is exalted to its utmost terms; on the one hand, it ensures the maximum guarantee; at the same time, it maintains a system that is based on a total and absolute rigidity of the penalty, a system that is therefore more 'limited'⁵⁴⁴.

On the contrary, in the second case, when one sees on the contrary the prevalence and the step backwards operated by the law; indeed, in the case of the absolute discretion entrusted to the judge, he is given the widest freedom of choice: both of the *an*, of the species of penalty but also of the *quantum*. The penalty system deriving from this model finds itself to be a system without predetermined limits; in fact, since it does not have rigid parameters to be adhered to pre-established by law, it gives rise to a very strong delegation for the identification of the criminal sanction⁵⁴⁵ which dominates the absolute indeterminateness of the penalty, to the detriment of the principle of legal certainty.

⁵⁴³ On this point, please refer to V. MILITELLO, *Prevenzione generale e commisurazione della pena*, Milan, 1982, 12 ss.

⁵⁴⁴ Suffice it to think, merely by way of example, of the absence of any form of individualisation of the sanction, which remains totally abstract with respect to the act committed by the offender. As is well known, this principle asserts itself above all in the 17th century. For more precise references, E. DOLCINI, *La commisurazione della pena. La pena detentiva*, Padua, 1968, 25 ss; V. MILITELLO, *Prevenzione generale e commisurazione della pena*, Milan, 1982, 8. The authors just quoted emphasise how, in this period, the principle is affirmed that the better the law, the less room it grants to the judge's discretion. In other words, it is intended to deny the judge any discretionary power in the application of punishment.

⁵⁴⁵ For instance, early Roman law is marked by the attribution of absolute discretionary power to the magistrate. Both the determination of punishable acts and the manner and forms of repression are left to his discretion. On

Lastly, then, placed in an intermediate position between the two previously illustrated, is the principle of constrained discretion, which is based, in theory, on the search for a point of equilibrium, certainly sacrificed by the previous guiding principles in the commensuration, between the guarantee requirement of absolute legality and the need for individualisation of the penalty.

Indeed, the systems that entrust this principle to the judge decide to delegate to the judge a certain amount of discretion, seeking, however, to direct it through certain limits and criteria predetermined by the legislature.

On closer inspection, after this brief and summary description of the principles that in a general way may influence the construction of the category of commensuration of punishment, the need to clarify the dogmatic concept of 'penal discretion' comes to the fore.

As far as the Italian legal system is concerned, the only explicit reference to discretion contained in the Criminal Code is to be found in the provisions of Articles 132 and 133 of the Criminal Code; indeed, these are provisions that regulate the discretionary power held by the judge in the phase of determining the penalty to be imposed in the specific case. This is the most important hypothesis of penal discretion that the legislator⁵⁴⁶ wanted to entrust to the judicial authority, which is called upon to adapt the sanctioning response to the peculiarities, objective and subjective, of the concrete case.

However, it is noted that in reality, in the Italian legal system, there are numerous provisions of the criminal code (and also of special legislation) which entrust, at least implicitly, in the area of sanctions, important judicial choices to the discretion of the judge⁵⁴⁷.

this point and in general on the evolution of the notion of commensuration of punishment in Roman law, U. BRASIELLO, (voce) *Diritto penale (diritto romano)*, in *Noviss. dig. it.*, Turin, 1964, 961 ss.; ID., (voce) *Pena (diritto romano)*, in *Noviss. dig. it.*, XIII, Turin 1965, 809 ss.; B. SANTALUCIA, (voce) *Pena criminale (diritto romano)*, in *Enc. dir.*, XXXII, Varese, 1982, 737 ss. In general, one can see an inversely proportional tendency of the affirmation of guaranteed principles with respect to the judge's discretion in the application of punishment. Whenever the weakness and inconsistency of the central power is affirmed, there is also an increase in the space, both in quantity and species, of arbitrary punishment.

⁵⁴⁶ On this point, M. SPASARI, *Appunti sulla discrezionalità del giudice penale*, in *Riv. it. dir. proc. pen.*, 1976, 53 ss.

⁵⁴⁷ For example, the main hypothesis is that concerning the commensuration of the penalty, governed not only by Article 132 of the Criminal Code for example, the main hypothesis is that concerning the commensuration of the penalty, governed not only by Article 132 of the Criminal Code, but also by Articles 133, 133 bis and 133 ter, the latter rules establishing a specific commensuration system for financial penalties; - the right to increase or decrease the penalty within very broad limits, if aggravating or mitigating circumstances are present; - again concerning aggravating and mitigating circumstances, the judgement of prevalence or equivalence within the institution of heterogeneous concurrence (Article 69 of the Criminal Code); - again, to remain within the same framework, the right to increase or decrease the penalty within very broad limits, if aggravating or mitigating circumstances are present. 69 of the Criminal Code); - again, to remain in the same field, the power to grant general mitigating circumstances (Article 62 bis); - the power to add a fine to imprisonment for offences committed for profit; - the power to reduce the sentence in the case of concurrence of persons in the offence with regard to those who have revealed a minimum capacity to commit offences, etc. This is obviously not an exhaustive list capable of

On closer inspection, what therefore stands out and makes, to some extent, the idea of a model using discretion and law wobble, are the numerous rules that allow and grant the judge wider spaces of power in the choice. Indeed, this tendency to considerably broaden the judge's discretionary powers, which is also found in some of the rules of the Code of Criminal Procedure and the Juvenile Judgement Code⁵⁴⁸, points to the fact that discretion in itself is a complicated terrain on which to move; moreover, it is extremely difficult to establish a priori, beyond the limits set by law, the boundaries of a choice that in itself has a subjective connotation.

For this reason, it was probably this field that, first and foremost in the United States, saw the emergence of instruments capable of supporting the judge in precisely this type of decision.

2.1.1 Initial reflections on the Italian discretionary system

On closer inspection, in the Italian system it is not possible to identify, either in the Criminal Code or in other normative sources, a ready and immediate definition of penal discretion⁵⁴⁹, since the only rules that mention it, Articles 132 and 133 of the Criminal Code, merely refer to this power but without delimiting its content, type or even function.

However, in order to maintain a more textualist reading, seeking to discern the legislature's original intention, a starting point for reflection could be the analysis of the positions taken by the doctrine together with the orientations of the jurisprudence of merit and legitimacy. Indeed, it cannot but be noted that one of the first doctrinal orientations to be formed on penal discretion is that which makes this concept coincide with an unlimited freedom of decision of the judge. Such a position seems very difficult to reconcile with the constitutional dictates on which the rule of law is founded; however, at the same time, it makes discretion coincide in all respects with mere arbitrariness, given that every time the legislator uses the adjective discretionary, he employs it in the sense of a synonym for 'optional', with the consequence that any hypothesis

encompassing all cases, but it does show how, in reality, beyond the model chosen by the Italian legal system, the same constrained discretion has various facets and boundaries. Moreover, one cannot fail to note that the matter of discretion in the application of punishment is also very wide-ranging within special legislation. And here too, for the mere purpose of giving a few examples, some of the powers entrusted to the supervisory magistracy are recalled, including - the power to apply the alternative sanction of probation to the social service (Art. 47 of the Penal Code); - the power to grant home detention (Art. 47 of the Penal Code) and semi-release (Art. 48 of the Penal Code).

⁵⁴⁸This can also be deduced from the same Report prefixed to the draft of the criminal code where we read. On the text of the aforementioned Report, see: *Lavori preparatori del Codice penale e del codice di procedura penale*, vol. V, *Progetto definitivo di un nuovo codice penale con la relazione del guardasigilli* on A. ROCCO, *Parte I*, *Relazione sul libro I del progetto*, Rome, 1929, VII.

⁵⁴⁹In this direction, F. BRICOLA, *La discrezionalità nel diritto penale*, vol. I, *Nozione ed aspetti costituzionali*, Milan, 1965, 9.

in which the exercise of discretionary power comes into play is thus referred to a totally free and unconditional decision and determination by the magistrate⁵⁵⁰.

What is to be noted is that such an orientation, in a perspective of comparison and contrast, is totally incompatible with the Italian constitutional order, which finds its foundations on the principles of the separation of powers and of taxation-legality in the criminal sector. In such terms, it appears almost obligatory to question the compatibility of the criminal rules and, in particular, of Article 132 of the Criminal Code, which admits and sets as the basic rule of the current penalty system the discretionary power of the criminal judge, with the guarantee system of the constitutional charter⁵⁵¹.

What is undoubtedly relevant, from a reading of the facts of the case, in relation to the phase of commensuration of the penalty in the strict sense, Article 133(2) of the Criminal Code requires the formulation of a criminological prognosis concerning the assessment of the offender's capacity to commit offences. Indeed, the judge, when he is called upon to identify the penalty to be imposed in concrete terms, must take into account (among other factors) the future conduct of the offender.

Once he has identified (ideally) the punishment proportionate to the offence committed, the judge may then impose a certain quantum of punishment below the proportion in order to satisfy special prevention requirements. Clearly, precisely such a finalistic orientation in the phase of commensuration of the penalty, finds its own constitutional foundation in the principle of re-education of the convicted person pursuant to Article 27⁵⁵², paragraph e of the Constitution.

⁵⁵⁰ On this P. NUVOLONE, *Corso di diritto penale, Parte generale*, Milan, 1966, 90, according to which this notion coincides with a form of discretion, which should be banned from modern criminal law. On this interpretative tendency, F. BRICOLA, *La discrezionalità nel diritto penale, vol. I, Nozione ed aspetti costituzionali*, Milan, 1965, 3; ID., *Il potere discrezionale del giudice nell'applicazione della sanzione criminale*, in *Monit. trib.*, 1968, 1229 ss. e E. DOLCINI, *La commisurazione della pena. La pena detentiva*, Padua, 1968, 59.

⁵⁵¹ In fact, the recognition in the head of the judge of the power in question, far from being in contrast with the principle of legality, is an expression, moreover, of other - and equally fundamental - constitutional principles. Among these are the principle of equality, which requires that different cases be treated differently, the principle of the personality of criminal liability, which requires that the penalty treatment be anchored to the degree of culpability, and the principle of the re-educative purpose of punishment, which cannot be achieved without careful consideration of the specific nature of each case. The argument was advanced by T. DELOGU, *Potere discrezionale del giudice e certezza del diritto*, 369; E. DOLCINI, *Note sui profili costituzionali della commisurazione della pena*, A. R. LATAGLIATA, *Problemi attuali della discrezionalità nel diritto penale*, in *Il Tommaso Natale*, Naples, 1975, 337 ss. Moreover, it was also the Constitutional Court itself that already in the 1970s emphasised, with reference to Article 25 paragraph 2 of the Constitution, and recalling even earlier precedents, that: <of the penalty (judgment no. 26 of 1966)>>. The Constitutional Court, 24 June 1970, no. 131, expressed itself in this sense. To compare the full text of the judgment see urlm <http://www.cortecostituzionale.it/ita/attivitacorte/pronunceemassime/pronunce>.

⁵⁵² In fact, there are also other provisions in the Code that refer to the need for a prognostic assessment. Again with reference to the moment of commensuration, Articles 102 et seq. of the Criminal Code implicitly require the judge to make a prognostic assessment if he intends to declare the convicted person a habitual, professional or

In conclusion, therefore, what has just been argued finds support in the very structure of penal discretion found in the Italian legal system; in fact, far from being a free and absolute discretion, it is a constrained discretion, whose boundary and margin of movement is delimited by the specification of limits to which the judge must adhere in determining the penalty⁵⁵³.

3 *Focus: the application of the sentence and the judge's discretionary power*

As is well noted, the choice of the Italian Code has been directed towards a penalty system based on constrained discretion.

In particular, in fact, the legal constraints imposed on the activity of commensuration of the penalty could be listed as follows: the edictal framework of the penalty that allows the judge to move within an edictal framework between a minimum and a maximum imposed *ex lege*; the explicit provision of commensuration indices provided for in Article 133 of the Criminal Code and, lastly, the obligation to state reasons provided for in Article 132 of the Criminal Code.

Therefore, in an attempt to delineate the picture of the Italian legal system more comprehensively, from a systematic point of view it is possible to distinguish two types of limits to legally binding discretion: on the one hand, there are the so-called 'internal' limits, which concern the very reason for the existence of discretionary power; on the other hand, there are the so-called 'external' limits, which instead coincide with those marked by all the rules circumscribing discretionary power⁵⁵⁴.

This distinction has actually been the basis of reasoning by other doctrine that identifies these limits: the internal ones, as edictal⁵⁵⁵, which coincide with the boundaries beyond and

trendy offender, after having determined the quantum of punishment; thus on this point, T. PADOVANI, *La pericolosità sociale sotto il profilo giuridico*, in Ferracuti (ed), *Psichiatria forense generale e penale*, Milan, 1990, 329.

⁵⁵³ In particular, the explanatory memorandum to the penal code specified that: "Now the parliamentary committee has asked whether this enumeration is peremptory or merely declaratory, expressing its preference for an indication that does not prevent the judge from taking into account other elements, such as anthropological data, illnesses prior to the offence, having committed the offence against the person to whom the offender owed gratitude, etc.". The enumeration is peremptory according to modern legislative technique, which has abandoned merely illustrative indications (of which there are a few examples in the criminal code of 1859 and also in that of 1889), because when a series of specifications is not peremptory, it is preferable to enunciate, instead of it, only the generic concept, into which the specific cases necessarily fall.

⁵⁵⁴ In particular, on this distinction by T. DELOGU, *Potere discrezionale del giudice e certezza del diritto*, 383; G. BELLAVISTA, *Il potere discrezionale nell'applicazione della pena*, 149 ss.

⁵⁵⁵ However, it should be noted that in order for the power conferred on the judge not to be transformed from a discretionary power into an arbitrary power, the edictal framework must not, in any case, range within excessively broad boundaries, the principle of legality requiring that "the breadth of the gap between the minimum and the maximum penalty must not exceed the margin of flexibility necessary to allow the individualisation of the penalty according to the criteria set forth in Article 133 and that it must be manifestly unrelated to the variability of the concrete cases and the types of subjects that can be related to the abstract case. 133 and which manifestly proves

below which the judge cannot go any further; the second ones, on the other hand, legally framed also as factual criteria (e.g. the gravity of the offence and the offender's capacity to commit a crime), guide the judge in the concrete determination of the penalty to be imposed; the judge is in any case required to take them into account in the statement of reasons, so as to allow a real control over the final decision.

Therefore, as regards the provision laid down in Article 133 of the Criminal Code, it can be seen from the outset that it is and represents the central provision of the penalty system, laying down the criteria that must be compulsorily followed by the judge in exercising his discretionary power⁵⁵⁶. By comparing this provision with the most relevant special legislation, it can be seen how in reality this provision constitutes and represents a directive criterion of general scope capable of guiding all the hypotheses in which the judge is endowed with margins of discretion⁵⁵⁷.

It should also be noted that a mere literal interpretation of the codified structure shows that albeit broad discretionary spaces present the character of so-called 'guided discretion'⁵⁵⁸.

The majority doctrine has taken a step forward by affirming that the listing is more all-encompassing than peremptory⁵⁵⁹.

What is certainly noticeable is that in the face of the vagueness and, in part, incompleteness of the data provided by the norms, one can see how the discretion bound on paper in reality corresponds to an 'empty' norm that, being also affected by the so-called 'non-choice' of the legislator on the function of the penalty, appears in part, also not very guiding.

to be unrelated to the variability of the concrete cases and the types of subjective cases that can be related to the abstract case. Otherwise, the legislative predetermination of the measure of the penalty would become merely apparent". Therefore, the very constraints placed on judicial discretion in the concrete determination of the penalty must find a reasonable limit in the law. 1 In this sense: Constitutional Court, 24 June 1992, no. 299, in *Giur. it.*, 1992, pp. 2033 ff. and in *Riv. it. dir. proc. pen.*, 1992, 1468.

⁵⁵⁶ According to part of the doctrine, Article 133 of the criminal code is an extremely general and vague provision and does not allow effective limits to be placed on the judge's discretion. In this sense, E. DOLCINI, *La commisurazione della pena. La pena detentiva*, Padua, 1968, 4

⁵⁵⁷ As, for example, one thinks merely by way of example of the application of security measures where the identification of one of the application prerequisites depends precisely on the criteria under Article 133 of the criminal code. Precisely for this purpose, Article 203(2) of the Criminal Code states that it is possible to infer the quality of a socially dangerous person from the circumstances set out in Article 133 of the Criminal Code. The examples in the code are numerous: think also of the definition of crimes of the same nature or the notions of habituality, professionalism or tendency to commit offences.

⁵⁵⁸ The legislator of 1930 did not consider it necessary to assess the dangerousness of the offender. In fact, the preface report to the code states: without listing the criteria required by Article 133 (on this point, please refer to the reading of the code), it is possible - for the sake of simplification - to summarise these elements in two general categories: the objective elements (referable to retribution), coinciding with the overall seriousness of the offence, and the subjective elements (referable to special prevention), i.e. the offender's capacity to commit crimes.

⁵⁵⁹ In this sense, F. BRICOLA, *La discrezionalità nel diritto penale*, which also quotes Massa's expression that: < there is no particular disposition of mind, personal state of the agent, quality or nuance of the action that cannot be included without dialectical effort among the elements of Article 133 c.p.

Therefore, the doctrine in the practice of application has endeavoured to identify how the indices of commensuration of the penalty operate. Therefore, in the light of the requirements of systematic classification, three types of criteria have been identified: finalistic, factual and logical⁵⁶⁰.

On closer inspection, as regards the first criteria mentioned, they consist in identifying the ends to be achieved by the imposition of the penalty. Indeed, the measure of the penalty and the type change depending on whether special, general or retributive prevention purposes are deemed prevalent⁵⁶¹.

Therefore, once the purposes to be pursued with the imposition of the penalty have been clarified, the judge must consequently select, in particular from Article 133 of the Criminal Code, only and exclusively those factual criteria that are ultimately compatible, i.e. relevant to the objective to be pursued with the penalty already identified in advance by the legal practitioner.

Lastly, the last stage of the procedure consists in assessing the different weight of the factual indices, taking into account that one is faced with an overall judgement that will thus lead one to choose the sanction in concrete terms between a predetermined *maximum* and *minimum*.

4 *American judicial practice and the use of risk assessment: a special focus on the selective incapacitation movement theory and evidence-based sentencing*

Artificial intelligence tools have provided solutions to the problem of the technical complexity, issues and costs of judicial operations, gradually penetrating even such a sensitive area of law. In particular, they have assumed in some countries (especially in US jurisdictions) a major weight in the sentenced person's judgement in assessing the risk of recidivism; in other countries, on the other hand, it has maintained an instrumental role in the judge's action.

In other countries, however, it has retained its instrumental role in the judge's action. On closer inspection, if one casts a glance at the American judicial practice, one can immediately find consolidated experience in the use of actuarial risk assessment instruments that support the judge in making decisions in the precautionary phase (the so-called *pre-trial decisions*), in the executive phase (*parole decisions*) and in the decisional phase (*sentencing decisions*)⁵⁶².

⁵⁶⁰ This systematic classification was proposed by E. DOLCINI, *La commisurazione della pena. La pena detentiva*, Padua, 1968, 4.

⁵⁶¹ Given the relevance of the purposes to be achieved by the penalty, it becomes pregnant - again for the purposes of commensuration - to also identify a hierarchy of the purposes of the penalty so as to guide the judge's work of commensuration of the penalty.

⁵⁶² G. CANZIO, *Il dubbio e la legge*, 4.

The aim of this paragraph and of this chapter introduction is to initially provide an overview of evidence-based sentencing in the context of the US criminal trial; in particular, the focus will be on algorithmic evaluation as a practice for assessing offender dangerousness.

This premise is considered most useful as it will allow the issue to be analysed from a comparative perspective with the Italian legal system; this will allow the possibility - while weighing the intrinsic difference between the two criminal and constitutional systems - to assess, from another point of view and from those who first anticipated this panorama, the first "perverse effects" and the first collisions derived from the application of these instruments with the guarantees of due process.

On the other hand, just as there are many critical points raised concerning the application of such instruments, at the same time, there are also considerable advantages that would seem to derive from their application: first and foremost, the demonstration of how the actuarial assessment of the offender's risk of reoffending is much more accurate than the human one, since it is able to process an immense amount of data that no judge could reasonably have at his disposal.

In fact, it is believed that the use of such assessment tools, also called mixed-type tools - since they are designed to give relevance not only to the statistical findings but also to the examination of the offender's personality - could offer a guide to direct the judge's activity in the commensuration of punishment⁵⁶³.

However, in practice and in reality, the boundary, its characteristics and its delimitation is not so strong or so marked.

In recent years, as already mentioned, several algorithmic risk assessment tools have made their way into various US civil service offices.

On closer inspection, the US courts⁵⁶⁴, in particular of Arizona, Colorado, Delaware, Kentucky, Louisiana, Oklahoma, Virginia, Washington and Wisconsin have shown a strong inclination to use risk assessment algorithms in recent years, to the extent that some scholars have even called it an '*Algorithmic Criminal Justice*'⁵⁶⁵. In fact, in recent years, risk assessment

⁵⁶³ L. D'AGOSTINO, *Gli algoritmi predittivi per la commisurazione della pena. A proposito dell'esperienza statunitense nel c.d. evidence-based sentencing*, 256.

⁵⁶⁴ There are more than 60 different types of risk assessment tools currently known, which take into account static and dynamic risk factors. To name a few LSI-R - Level of Service Inventory - Revised ,LSI/CMI - Level of Service/Case Management Inventory , ORAS - Ohio Risk Assessment System ,Static-99 (for sex offenders/offenses only), STRONG - Static Risk and Offender Needs Guide ,Wisconsin State Risk Assessment Instrument.

⁵⁶⁵ «Algorithmic criminal justice, as I define the term, is the application of an automated protocol to a large volume of data to classify new subjects in terms of the probability of expected criminal activity and in relation to the application of state coercion»; A.Z. HUQ, *Racial Equity in Algorithmic Criminal Justice*, in *Duke Law Journal*, 2019, 1060.

tools have been recognised as the key instrument of the criminal justice bail reform that has affected the United States.

Indeed, it is precisely the collection and processing of vast amounts of data (big data) that has been welcomed by national governments in awareness of the incredible potential they offer. As already mentioned, the trust placed in such tools emerges even more clearly from their use in various areas of criminal justice: from the now diverse and considerably implemented predictive policing techniques, to the assessment of the offender's social dangerousness⁵⁶⁶.

On closer inspection, the application of machine learning and crime prevention tools, above all, represents, according to many scholars, the normal evolution of an earlier trend towards the use of risk assessment tools based on statistical-actuarial calculations

On closer inspection, the application of machine learning and crime prevention tools above all represents, according to many scholars, the normal evolution of an earlier trend towards the use of risk assessment tools based on statistical-actuarial calculations.

Therefore, before turning our attention and focusing on modern risk assessment software, it would first appear useful to recall the historical-dogmatic foundation underlying the use of such tools.

In fact, according to some authors, the modern debate on the algorithmic assessment of risk-offence presents strong similarities with that which arose concerning the theory of selective incapacitation (the so-called *selective incapacitation movement*)⁵⁶⁷.

It is precisely this theory that starts from the premise that the criminal justice system should be conformed to allow a precise identification of socially dangerous categories of subjects - or prone to violence or professional or tendency criminals - so that they can be neutralised by keeping them in prison for long periods of time: indeed, what one would like to achieve is that the 'elimination of such subjects from society' would lead to an overall reduction in the crime rate⁵⁶⁸.

⁵⁶⁶ This expression generally refers to the set of methods and techniques used by the public security authorities to prevent the commission of crimes. Recently, the subject has been taken up again in connection with the use of predictive algorithms to indicate to the police in real time, according to probabilistic criteria, the metropolitan areas to be controlled or guarded. L. BENNET MOSES - J. CHAN, *Algorithmic Prediction in Policing: Assumptions, Evaluation, and Accountability*, 806

⁵⁶⁷ On this subject, see the article by the Harvard Law Review Association, *Selective Incapacitation: Reducing Crime Through Predictions of Recidivism*, in *Harvard Law Review*, 1982, 96, 2, 511 ss.

⁵⁶⁸ Modern risk assessment algorithms are programmed to express a judgement of dangerousness by processing data on categories of subjects distinguished by age, lifestyle, family composition, origin, etc. that can guide judges in determining the punishment to be imposed on the concrete case. It may be noted that the starting premises and the decisional outcome (more severe penalty for an individual deemed socially dangerous due to belonging to a 'category') faithfully reflect the postulates of the selective incapacitation theory.

Indeed, it must be remembered that crime prevention through the instrument of 'prediction' has accompanied the US criminal justice system since the 1920s; only later - around the 1970s - did studies focus mostly on the search for indices of dangerousness that would confirm the subject's aptitude to commit violent crimes⁵⁶⁹.

What principally emerged from the first studies was undoubtedly the difficulty of being able to establish and trace objectively the revelatory indices of current dangerousness; indeed, precisely in this sense, the advocates of the theory proposed, following a utilitarian approach, to punish certain individuals more severely on the basis of the sole positive prognostic judgement of recidivism in crime⁵⁷⁰.

However, the method's lack of scientific reliability has also contributed to making the theory of selective incapacitation a relic of the past; however, there remain some institutions that still seem to recall it⁵⁷¹.

In conclusion, it seems only useful to reflect that according to some scholars, this type of statistically-based assessment of the dangerousness of the offender is the product of the legal-philosophical reconstruction of the US doctrine on the objectives of criminal justice. In particular, the Enlightenment theory of the re-educative function of punishment, the so-called rehabilitation, should be credited with the principle according to which a sanctioning treatment that enhances the characteristics of the individual rather than the offence he has caused should be privileged; indeed, the predetermined edictal penalties would constitute an obstacle to an individualised punishment. On the contrary, the judicial discretion in the commensuration of the penalty, although it looks at and is functional to the re-education of the convicted person, had however, at the same time, raised several questions on the level of equality of treatment⁵⁷².

⁵⁶⁹ However, the prediction of dangerousness proved to be quite complex and the first attempts resulted in a considerable number of false positives. J. COHEN, *Incapacitation as a strategy for crime control: possibilities and pitfalls*, 1-84 in M. Tonry and N. Morris (eds.), *Crime and Justice: An Annual Review of Research*, Vol. 5. Chicago, 1983, 12.

⁵⁷⁰ The theory was based on the assumption that professional or trend criminals - responsible for the most serious crimes - can be easily identified from certain known characteristics, such as their personal and criminal history. However, the choice of punishing criminals not for what they had already committed, but for what they might commit in the future, clashed with the argument of those who, on the basis of statistical evidence, showed that the expected crimes might in practice never be committed. T. MATHIESEN, *Selective incapacitation revisited*, in *Law Human Behaviour*, 22, (4) 1998, 455.

⁵⁷¹ Several States have introduced autonomous discipline statutes for serial offenders, established specialised departments in public prosecutors' offices for proceedings against professional criminals, and required judges to take into account criminal records, employment stability and other personal data.

⁵⁷² However, leaving too much discretion in the hands of judges could have negative effects in terms of equal treatment. In the course of history, it has in fact happened that convicted persons belonging to social minorities (linguistic, ethnic, racial) have suffered disproportionate treatment compared to members of the dominant or high social classes. D. KEHL – P. GUO – S. KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing, Responsive Communities Initiative*, 6.

4.1 *Segue: the penalty phase in the US system*

As anticipated, especially in the United States, the debate on the individualisation of punishment has favoured a shift towards indeterminate sentences, assigning the discretionary decision on the duration of deprivation of liberty to expert commissions rather than to the judge⁵⁷³. This trend, theorised since the 1870s, was originally not generally accepted. Indeed, it was strongly criticised for its incompatibility with the traditional principles of *nulla poena sine lege and punitur quia peccatum est*⁵⁷⁴. On the one hand, the traditional balance of powers, legislative, judicial and administrative, was at stake: not only does the indeterminate sentencing system remove the legislature's power to set minimum and maximum sentences (as provided for in the Federal Constitution), but it also severely limits the judge's discretion in sentencing, which is reserved to him, replacing it with the (possible) arbitrariness of a purely administrative body not bound by the general principles of criminal law⁵⁷⁵. On the other hand, the implicit foundation of all modern Western ideology, i.e. free will, seemed to be thrown into doubt: criminal responsibility and punishment lost their classical centre of gravity, i.e. culpability (i.e. the individual's reproach for his free choice to act against the criminal law), ending up projected into the sphere of 'criminal law of the enemy', based on the character and criminological type of the perpetrator⁵⁷⁶.

On closer inspection, the advocates of 'individualised' and, therefore, indeterminate punishment prevailed because of the strong and generalised dissatisfaction with the criminal justice system of the time, considered as a whole. The early years of the 20th century saw a growing feeling of strong frustration, due to the obvious inequalities between the penal systems and criminal policy strategies of the different federal states. Legal historians report enormous concern about the biases and arbitrariness of state courts, which were poorly guided and constrained in sentencing by local norms. Faced with a multifaceted, incoherent, hardly harmonisable context, the proposal - supported by the theories of the emerging psycho-criminological science - to assign the task of sentencing to expert commissions, operating on the basis of a common index, i.e. science, met with success. Moreover, in the opinion of many, the execution of the (indeterminate) sentence and the periodic re-evaluation of its outcome

⁵⁷³ See, S. QUATTROCOLO, *Sui rapporti tra pena, prevenzione del reato e prova nell'era dei modelli computazionali psico-criminologici*, in *Teoria e critica della regolazione sociale*.

⁵⁷⁴ *Ibidem*.

⁵⁷⁵ See, however, with regard precisely to the Italian legal system, the reflections of E. DOLCINI, *La commisurazione della pena. La pena detentiva*, 55, on the capacity of Articles 132 and 133 of the Code of Criminal Procedure, as interpreted by case law.

⁵⁷⁶ Risk expressly reported, with regard to digital risk assessment, da G. UBERTIS, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, in *Sistema penale*, 10.

were to be considered, more correctly, functions of an administrative, rather than judicial, nature, to be assigned to a non-judicial body, such as the prison board⁵⁷⁷. Thanks to the bifurcation between verdict of guilt and sentencing, the advocates of the theory of indeterminate punishment were able to advocate a division of punitive power between judicial and administrative bodies that did not require constitutional amendments, with the result that the widespread fear of an 'administrativeisation' of criminal justice was easily overcome: this resulted, as early as the 1920s, in the general spread, in the American states, of indeterminate sentencing systems, based on individualised and progressive treatment of prisoners and administered by 'committees of experts'. Hence a profound distinction between the United States and the continental European tradition, where, as mentioned, the bifurcation between verdict on fact-finding and sentencing was, and still is, rare.

In light of this situation, the 20th century marked, in the United States, the rise of 'predictive sentencing' theories, traditionally based on the prediction of the offender's future behaviour, fuelling psycho-criminological research on the risk of violent and recidivist behaviour. The results of that research, in fact, pushed towards a specific paradigm: risk assessment and its prevention have always been associated with the incapacitation of the convicted person, i.e. with the deprivation of personal liberty until his dangerousness ceases. However, as already mentioned, risk assessment, based on criminogenic factors, i.e. risk assessment, is a tool for measuring antisocial behaviour, not a response to it. Despite this clear evolutionary line, the debate on the most appropriate balance between retribution and prevention in sentencing has never been exhausted. The recent history of the penological debate in the United States highlights the existence of oscillating trends. In the 1970s a revival of retributivist theories, based on proportionality to the fact of crime, began, a revival inspired by the massive incarceration rates that revealed (and still reveal) a severe impact of prison sentences on minorities and the poorer strata of society. These findings suggested, according to many, the need to move from an individualised sentencing system to strict compliance with federal sentencing guidelines, a useful harmonisation tool to accompany the guidelines developed in the various states. Already in the following decade, however, the United States saw a new trend advancing towards the revival of the so-called 'selective incarceration' of individuals deemed most dangerous, based on the assumption that 'career criminals' can be identified through their personal and psychological characteristics and criminal history and, therefore, effectively segregated. Despite serious doubts as to the accuracy of the underlying psycho-criminological

⁵⁷⁷ On this point, the decision *Woods v. State*, 130 Tenn. 100, 114, 1914.

theory, the doctrine of selective incapacitation was so successful that it greatly influenced many North American jurisdictions. In particular, in the 1990s, the school of the so-called 'new penology' succeeded in affirming 'social risk management' as the main function of punishment, to the detriment of rehabilitation, confirming the tendency to link the quantification of punishment to the outcomes of risk assessment. It is precisely the recent digitalisation of risk assessment tools that has significantly revived the long-standing dilemmas mentioned above⁵⁷⁸.

Is it acceptable to link the quantification of the sentence to the assessment of the risk of reoffending? How accurate and reliable are the premises and results of this risk assessment? Is incapacitation, resulting from the extension of the indeterminate sentence, the only possible way to deal with the social dangerousness of the offender? The passage of time and the change in society have not reduced the variety of arguments involved in the debate. Indeed, the spectrum of positions is broad and ranges from the affirmation of the full compatibility of indeterminate punishment with the principle of guilt, to the denunciation of the unacceptability, empirical and moral, of a punitive system based on the prediction of the risk of dangerousness, calling for its abandonment. And on these profiles, the present reconstruction cannot offer any useful contribution. However, the new digital turn of risk assessment tools has added a further level of complexity to the topic in question, and an attempt will be made below to trace a systematic framework.

4.2 *How actuarial risk assessment came about*

The penal landscape in the United States changed considerably with the 1984 reform law that had helped mark the transition from the rehabilitative to the retributive conception of punishment. Indeed, through this reforming intervention, the idea had been accepted that criminal sentences should be commensurate with the extent of the offence and the damaging consequences and, at the same time, weighted on the basis of the elements identified in the best practices most widely used at the federal level⁵⁷⁹.

Indeed, the main problem that ensued was that this new model that was emerging risked leading to the mirror-image situation: in fact, if the discretion granted for rehabilitation gave rise to unequal treatment, at the same time the total compression of the judge's discretionary power led to results contrary to the requirements of substantive justice. As a consequence, the

⁵⁷⁸ K. HANNAH-MOFFAT, *Unpacking sentencing algorithms: Risk, racial accountability*, 270; D. KEHL – P. GUO – S. KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing, Responsive Communities Initiative*, 6.

⁵⁷⁹ *Ibidem*.

problem of prison overcrowding (which for many was the derivative of the "massive expiation of short prison sentences imposed in the aftermath of the sentencing reform") was presented and increased.⁵⁸⁰).

And it is precisely in this context⁵⁸¹ that the conviction began to spread that judges should base their decision on the amount of the sentence and, where appropriate, that on the granting of reward benefits or alternative measures to detention, also based on statistical evidence, the so-called evidence-based practices⁵⁸².

Thus, the actuarial assessment of the risk of recidivism in the offence would have enabled the judge to make more informed determinations and to choose the coercive measure or the quantity of punishment most appropriate to the case⁵⁸³. The basic idea was indeed based on the conviction that decision making in the criminal justice process could no longer do without scientific knowledge⁵⁸⁴. However, according to the prevailing opinion, such an approach risked then constituting the perfect synthesis between the rehabilitative paradigm and the retributive paradigm: this is also explained by the fact that the judge in this case would be bound to take into account the objective elements of the fact, without neglecting the factors relating to the personality of the offender and his aptitude to commit offences. In fact, originally, risk assessment was carried out on a case-by-case basis by prison psychologists, who relied on their professional knowledge and on the results of the offender's rehabilitation process on a case-by-case basis. However, this system had several shortcomings, first and foremost the difficulty of expressing measurable and comparable results and, moreover, being almost completely unusable in the judicial phases preceding the execution of the sentence.

⁵⁸⁰ L. D'AGOSTINO, *Gli algoritmi predittivi per la commisurazione della pena. A proposito dell'esperienza statunitense nel c.d. evidence-based sentencing*, 256.

⁵⁸¹ According to the prevailing opinion, such an approach would constitute the perfect synthesis between the rehabilitative and retributive paradigms. The judge would in fact be bound to take into account the objective elements of the fact, without neglecting the factors relating to the offender's personality and his aptitude to commit offences. Originally, the risk assessment was carried out on a case-by-case basis by prison psychologists, who relied on their professional knowledge and the results of the offender's rehabilitation. This system had the defect of expressing results that were difficult to measure and compare with each other, as well as being unusable in the judicial phases preceding the execution of the sentence. Over the years, evidence-based-practise has been supported by increasingly sophisticated predictive tools that consider the interaction between static and dynamic risk factors. Next-generation tools use machine-learning algorithms.

⁵⁸² Recently on this issue, M. KINGELE (et oths), *Prognostic value of procalcitonin in patients after elective cardiac surgery: a prospective, cohort study*, in *National Library of Medicine*, 2016, 537.

⁵⁸³ Evidence-based practices use data on socio-economic conditions and the results of specific tests to assess the dangerousness of the offender and the risk of reoffending; the aim of these methods is to reduce the likelihood of re-offending. Offenders are generally grouped according to their score into three risk bands (high, medium and low).

⁵⁸⁴ «As in medicine, psychology, education, management, and other fields, science now offers empirically-derived practice guidelines for criminal justice, which is part of a gradual trend towards the use of evidence-based practices in law». R. E. REDDING, *Evidence-Based Sentencing: The Science of Sentencing Policy and Practice*, in *Legal Studies Research Paper Series*, Paper No. 09-41, 2.

Over the years, evidence-based practice has certainly been able to rely on increasingly sophisticated predictive systems and tools that consider the iteration between static and dynamic risk factors. Indeed, new-generation tools use machine learning algorithms capable of weighting these factors by processing immense amounts of data⁵⁸⁵.

4.3 *Compas: the Loomis case and the Wisconsin Supreme Court decision*

On closer inspection, one of the cases that caused the greatest stir and focused and channelled the debate on risk assessment tools used in the sentencing phase was the Loomis case of 2016. This case served to highlight the first critical issues arising from the application of a risk assessment and hazard identification tool called COMPAS⁵⁸⁶.

In fact, this tool immediately showed critical elements linked, above all, to its actual predictive validity (*accuracy*) and its impartiality (*fairness*)⁵⁸⁷.

In the Loomis case⁵⁸⁸, the Wisconsin Supreme Court, however, denied that the defendant's inability to assess the scientific reliability of COMPAS, due to its secrecy, caused an infringement of due process⁵⁸⁹: the Court held that the defendant could, on the basis of the instrument's user manual, compare individual data (i.e. input) and final risk assessments (output), thus refuting reliability⁵⁹⁰. Moreover, the use of COMPAS was only considered

⁵⁸⁵ Currently, the laws of many states provide that the courts can - and in many cases must - consider the output provided by the algorithm before making a decision.

⁵⁸⁶ The factors that COMPAS takes into account are: COMPAS takes into account - in its basic configuration - the answer to 137 questions, concerning the following items: - criminal history; - previous misdemeanours and offences; - past violence; - current violence; - acquaintances with criminals; - substance abuse; - economic problems; - difficulties in education and vocational training; - delinquent family environment; - social context; - way of using leisure time; - residential instability; - social adjustment; - socialisation defects; - criminal opportunities; - social isolation; - pro-criminal thinking; - criminal personality.

⁵⁸⁷ On this point, F. BASILE., *Intelligenza artificiale e diritto penale*, 17.

⁵⁸⁸ For a brief review of the case involving the defendant Eric Loomis, who was involved in a shooting, the latter had pleaded guilty to two of the five counts (driving a vehicle without the owner's consent and attempted violation of a roadblock) and the local court had sentenced him to six years' imprisonment and five years' extended supervision, basing its decision, at least in part, on a 'high risk' prediction provided by COMPAS. Following the rejection of a petition for post-conviction release, the defendant appealed to the Supreme Court, claiming, firstly, violation of the defendant's right to be assessed on the basis of accurate information; secondly, violation of the right to an individualised sentence and, finally, male gender among the various data used to assess dangerousness. For an accurate reconstruction, see S. QUATTROCOLO, *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs. rischi e paure della giustizia digitale 'predittiva'*, in *Cass. Pen.*, 2019. The case caused quite a stir in public opinion and, in the general vernacular, became one of the paradigmatic examples of the substitution of machine for man: see, A. LIPTAK, *Sent to Prison by a Software Program's Secret Algorithms*, in *The New York Times*, 1st March, 2017.

⁵⁸⁹ The Supreme Court of Wisconsin, prompted by these findings, issued a warning in relation to the future use of COMPAS, highlighting: - its nature as a product covered by industrial secrecy, which prevents the disclosure of information relating to its method of operation; - the fact that the evaluations are carried out by COMPAS on a collective, group, and not individual basis; - finally, the risk of overestimating the risk of crimes being committed against certain ethnic minorities.

⁵⁹⁰ *State v. Loomis*, 881 NW 2d 749 (Wis 2016), § 53-54. For a commentary on the judgment See Criminal Law – Sentencing Guidelines – Wisconsin Supreme Court Requires Warnings before Use of Algorithmic Risk

legitimate in the presence of certain countervailing factors: firstly, the Court stated that "a circuit court must explain the factors in addition to a COMPAS risk assessment that independently support the sentence imposed. A COMPAS risk assessment is only one of many factors that may be considered and weighed at sentencing". Secondly, the Court required that five warnings be given to the judge in the Presentence Investigation Report ('PSI'), among which it is worth noting the second one, according to which, since the risk assessment is based on data referring to classes of subjects, the COMPAS⁵⁹¹ is able to identify groups of persons at high risk of reoffending and not a single high-risk individual.

However, where such conditions exist, the Court considers that "consideration of a COMPAS risk assessment at sentencing along with other supporting factors is helpful in providing the sentencing court with as much information as possible in order to arrive at an individualised sentence".

On this occasion, the position of the US Supreme Court of Wisconsin was strong and clear: it upheld this decision and rejected the writ of certiorari filed against it. This pronouncement would seem to point in the direction of an openness and acceptance of such tools; the literature and the reflections that were then triggered in turn endorse the position of those who believe that the creation of an algorithm, carried out in an appropriate manner, can go towards perfecting man's predictive decisions, which are naturally based on limited experience: the tools would in fact contribute to reducing the prison population and ensuring the elimination of racial disparities, thus becoming 'a force for racial equity'.

However, in this case, the Supreme Court has, at the same time, drawn up a sort of 'cautionary decalogue'⁵⁹² which judges must employ in the use of such 'predictive' tools, articulated in five warnings that must always be included in the pre-sentencing report, namely the possible existence of a trade secret covering the software; the inability of the software to

Assessment in Sentencing – *State v. Loomis*, in *Harvard Law Review*, 2017, 1530 ss. The use of COMPAS in sentencing had already been admitted in Wisconsin by the ruling *State v. Samsa*, 2015 WI App 6.

⁵⁹¹ To get an idea of how COMPAS works, one can browse online through a version of its 'Operations Manual', dating from March 2015, from which one learns, among other things, that: "COMPAS is a fourth-generation [offence commission] risk and [treatment] needs assessment tool. Criminal justice agencies across the country use COMPAS to make decisions about the placement, supervision and management of offenders. COMPAS was developed empirically with a focus on predictors known to influence recidivism. It also takes into account dynamic risk factors and provides information on a variety of widely validated [by scientific research] risk factors in order to facilitate corrective interventions aimed at reducing the likelihood of reoffending [...]. COMPAS was first developed in 1998 and has since been revised several times as the knowledge base provided by criminology and correctional practice has evolved [...]. We continue to make improvements to COMPAS based on the results of empirical research and recidivism studies conducted in prisons or probation agencies. COMPAS is periodically updated to keep up with emerging best practices and technological advances [...]"

⁵⁹² Così, S. QUATTROCCOLO, *Sui rapporti tra pena, prevenzione del reato e prova nell'era dei modelli computazionali psico-criminologici*, 273.

make a highly individualised assessment, being based on a set of data referring to social groups, not normalised with respect to the population of each State; the creation of the tool for purposes specifically related to choices proper to the executive phase, subsequent to sentencing, as well as the existence of doubts, in the scientific community, as to the reliability of the computational model - albeit secret - that regulates it.

As the other side of the coin, critically, the Wisconsin Supreme Court is considered to have set a 'dangerous' precedent by opening the door to the widespread use of COMPAS in the courts and also affirmed that the right of access to the algorithm is not granted to the defendant but that the defendant's right to a fair trial is not impaired.

In other words, as a corrective to avoid abuse in the use of these tools, the Court reaffirmed that they have a merely instrumental and functional role in identifying the specific needs of the defendant.

4.4 The peculiarity of the decision: the decisional 'double phase' in the choice of penalty treatment

In order to fully explain and understand part of the reasons why it has been easier to introduce these instruments within the judge's difficult decision-making process, it is first appropriate to assess the peculiarities and characteristics of this phase within the North American panorama.

In fact, the procedure followed by the District Courts in the USA to determine the punishment of a convicted person is characterised by a procedure that is divided into two phases: in fact, in a preparatory or preliminary phase, an attempt is made to obtain, through a study of the defendant, his socio-criminological profile⁵⁹³; only at a later stage, when the trial is closed and the sentence is pronounced, the judge must wait to receive the presentencing investigation report (the so-called PSI) in which the results of the investigation are presented to the court. so-called PSI) in which elements are included that will be useful in determining the type and amount of the sentence.

This report is prepared by an auxiliary (in most cases) who has considerable experience in the social welfare sector or expertise in the psychological or criminological sciences. On closer inspection, the report includes other information concerning the characteristics of the subject who is subjected to this report: there are in fact details of his past life, such as his biography, criminal record, findings from family interviews or interviews with former employers, friends

⁵⁹³ L. D'AGOSTINO, *Gli algoritmi predittivi per la commisurazione della pena. A proposito dell'esperienza statunitense nel c.d. evidence-based sentencing*, 360 s.

and associates. Perhaps the most peculiar and important aspect is that there are no limits set by law on the elements that can be included in the report. In fact, it would almost be a sort of 'special' investigation that is removed from the general principles of the accusatory model⁵⁹⁴; it follows that the judge is free to take into consideration all the elements he deems useful and even if they do not appear to have been the subject of cross-examination between the parties.

Once it is then deposited in the clerk's office of the district court, this report will be available and accessible to the defence, which may examine it⁵⁹⁵ (with certain exceptions).

The pre-trial phase preceding the imposition of the sentence is then the phase of the trial in which algorithmic risk assessment has become most prominent.

Indeed, looking at the most recent legislation, some scholars have also noted that there is a growing tendency over time for states to impose increasingly stringent sentencing constraints on judges, which would seem to increasingly compress the sphere of discretion entrusted to the sentencing body⁵⁹⁶.

In conclusion, to date, A.I. tools are applied in most states in the USA: in some jurisdictions, others, the use of such risk assessment tools is even mandated by law. For example, in Arizona, PSI is specifically required to contain specific information 'related to criminogenic risk and needs as documented by the standardised risk assessment and other file and collateral information'⁵⁹⁷.

Likewise, Oklahoma requires the use of 'assessment and evaluation instruments designed to predict risk of recidivism to determine eligibility for any community punishment'⁵⁹⁸.

A law of the State of Ohio had entrusted the Department of Correctional Justice with the task of identifying a reliable risk assessment instrument that could be used for various purposes, including the commensuration of punishment. The Ohio Risk Assessment System (ORAS), a

⁵⁹⁴ The exception is justified by the peculiarities of this procedural phase. The investigation into the personality of the defendant could certainly not be conducted before or during the trial, for obvious reasons of extraneousness to the *thema probandum* and respect for the principle of the judge's impartiality and third party *status*.

⁵⁹⁵ However, access to certain parts of the document or to certain information classified as confidential may be restricted. Restricting access ensures that individuals heard during the investigation are protected from possible retaliation by the convicted person, thereby incentivising them to cooperate with justice. Once the PSI has been filed, the trial ends with the final so-called sentencing hearing, at the end of which the judge will take - following his or her free conviction - the decision on the penalty to be imposed on the convicted person, based on all the available evidence, including the evidence that the parties present at the same hearing.

⁵⁹⁶ D. KEHL – P. GUO – S. KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing, Responsive Communities Initiative*, 15 to which reference is made for more details on the legislation of individual states on the use of algorithmic and actuarial risk assessment tools. <https://epic.org/algorithmic-transparency/crim-justice>.

⁵⁹⁷ *Arizona Justice Administration Code*, § 6–201, 01(J)(3).

⁵⁹⁸ D. KEHL – P. GUO – S. KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing, Responsive Communities Initiative*, 15.

risk assessment software developed by a team of experts and academics at the University of Cincinnati, was thus created⁵⁹⁹.

There are also different situations in other states⁶⁰⁰ where a more 'cautious' approach and attitude has been adopted: instead of introducing such tools head-on, making them mandatory for judging bodies, it has been decided to promote so-called good practices (or even BSE) without, however, imposing the mandatory use of predictive tools or algorithms⁶⁰¹.

Indeed, the panorama that presents itself is very diverse, mostly due to the fact that this is a subject that has neither common nor homogeneous regulations in the different jurisdictions.

For this reason, a proposal has recently been made to amend the Moal Penal Code, which would in part amend some of the sentencing provisions that deal precisely with risk assessment tools. In this future perspective, it would be envisaged that judges should consider the results of risk measurement before sentencing, since statistical-actuarial assessments, derived from objective criteria, that have been found superior to clinical predictions built on the professional training, experience, and judgment of the persons making predictions. In short, recidivism risk prediction is inevitably part of sentencing, and rather than being guided by judges' unreliable 'clinical' assessments of offenders, it should be guided by the best available scientific research".

What is undoubtedly emphasised and on which the greatest possible⁶⁰² leverage is sought is the need to ensure the accuracy and reliability of these instruments, in order to ensure that they are then used in a transparent manner and with respect for the offender's right of defence⁶⁰³.

However, as in every subject, there are also voices to the contrary. In fact, the Department of Justice has expressed a rather sceptical stance towards predictive algorithms, warning national legislators about the possible discriminatory effects of their use on individuals from disadvantaged social classes⁶⁰⁴.

⁵⁹⁹ *Ohio Revised Code*, § 5120.114(A), (1-3).

⁶⁰⁰ Among them, Louisiana, Idaho, Indiana, Maryland, Alaska.

⁶⁰¹ In a recent case, the Indiana Supreme Court urged judges on the merits to make use of such tools, emphasising emphatically that the scientific literature «has demonstrated for decades that objective actuarial risk/needs instruments more accurately predict risk and identify criminogenic needs than the clinical judgment of officers». Thus, *Malenchik v. State*, sentenza del 09 giugno 2010, repertorio dello Stato dell'indiana n. 928 N.E.2d 564, § 7.

⁶⁰² In recent times, the highest institutions of the US judicial system, including the Conference of Chief Justices and the Conference of State Court Administrators, have launched a number of initiatives to develop best practices for evidence-based sentencing. Among the goals of the project is to reduce the rate of prison sentences to be served through accurate profiling of offenders at low risk of reoffending.

⁶⁰³ In argomento, S. B., STARR, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, in *HeinOnline*, 66 Stan. L. Rev., 2014, 815.

⁶⁰⁴ D. KEHL – P. GUO – S. KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, *Responsive Communities Initiative*, 16.

4.5 *An application overview of the United States: the case of Virginia*

On closer inspection, the State of Virginia was the first state to implement a risk assessment instrument for use to be used in sentencing phase. The instrument was created by the Virginia⁶⁰⁵ in 1994 Criminal Sentencing Commission and was designed, initially, to identify low-risk felons in order to assign them a more suitable type of punishment⁶⁰⁶.

These alternative punishments include diversion from prison to jail, diversion from jail to community service or home-arrest, and fines.

It seems appropriate to note, in a continental landscape that has seen the introduction and then the subsequent introduction of such instruments in the criminal justice system, that the State of Virginia, from the very first moment, remains unique in “its approach to developing risk assessment tools”. Indeed, “while a handful of states, like Virginia and Pennsylvania use risk assessment tools that have been developed by (or In partnership with) the state government, many more States and jurisdictions have implemented or adapted one of several existing commercial systems”⁶⁰⁷.

On closer inspection, one of the earliest and most popular risk assessment tools in sentencing was called the Level of Service Inventory-Revised (LASI-R)⁶⁰⁸. This tool, which was developed by the Canadian company Multi-Health Systems, pulls information from a survey containing a wide set of static and dynamic factors. Furthermore, these factors, which range from criminal history to personality patterns, are used to determine a person’s risk for recidivism as well as the best sentencing options. Indeed, the tool was initially developed to be used in rehabilitations, however it was later adopted for use in the sentencing phase⁶⁰⁹.

Another tool that only needs to be mentioned is COMPAS. This tool assesses variables under five main areas: criminal involvement, relationships/lifestyles, personality/attitudes, family and social exclusion. It uses a combination of static and dynamic factors in order to assess recidivism risk, and it can be programmed for a variety of use cases. Although Compass

⁶⁰⁵ *Ibidem*, 11.

⁶⁰⁶ B. OSTROM, *Offender risk assessment in Virginia: A three-stage Evaluation: process of sentencing reform, empirical study of diversion and recidivism, Benefit-cost analysis*, National Center for State Courts: Virginia Criminal Sentencing Commission, 2022.

⁶⁰⁷ Algorithms in the Criminal Justice System, Electronic Privacy of information center.

⁶⁰⁸ The Virginia Pretrial Risk Assessment Instrument (VPRAI) is used in 16 counties and 1 state. Another particularly innovative programme promoted by the Virginia Criminal Sentencing Commission, the Nonviolent Risk Assessment (NVRA), is also used in Virginia to identify low-risk drug offenders and property offenders for application of alternative measures to imprisonment.

⁶⁰⁹ LSI-R and adapted versions of it are used to assist sentencing in a number of states and jurisdictions, as for example, Washington and California. See, Sex offender sentencing in Washington state: predicting recidivism based on the LSI-R, Washington state institute for public policy, 2006.

can be employed for purposes beyond sentencing, a number of states, including Wisconsin, Florida, Michigan, use COMPAS in order to assist judges with sentencing decisions⁶¹⁰.

5 *From risk assessment in recidivism to sentence commensuration: why algorithms fit into sentencing*

On closer inspection, predictive algorithms were initially applied in the pre-trial phase because their importance was recognised for those evaluations that already in themselves implied a prognostic assessment looking to the future.

It is no coincidence, in fact, that one of the main issues addressed by American doctrine concerned precisely the relationship between the risk of recidivism and the commensuration of punishment. Indeed, it was precisely the fact that predictive algorithms were already being used in other phases of the criminal trial that had led to the need to try to broaden the field of application, coming to find application also in the sentencing phase.

Moreover, it is no coincidence that the assimilation of the two decision-making contexts should follow a certain caution. Indeed, in most cases, software is programmed to support decisions in the pre-trial release phase; in fact, the prognostic judgement that is made at this stage is intended to predict whether or not the defendant will refrain from committing other crimes⁶¹¹ during the trial (it will therefore have to assess whether or not there are any pre-trial needs).

On the contrary, when a conviction is pronounced, the decision-making process will be much more complex and articulated, since the judge will have to determine on the one hand the punishment to be imposed and, on the other hand, also the choice of measure.

In this regard, it is noted how the determinations on the *quantum puniatur* are conditioned by the different theories on the functions of punishment (retributive, re-educative, preventive). Part of the debate and of the central themes that will touch this chapter, albeit in a transversal manner, is centred on this point; on closer inspection, there are some scholars⁶¹² who believe

⁶¹⁰ However, there are not a few criticisms and issues that have arisen and been generated as a result of the first applications of COMPAS. Indeed, since COMPAS is proprietary software, “it is not subject to federal oversight and there is almost not transparency about its inner working, including how it weights certain variables. COMPAS has created a considerable amount of controversy for this very reason”. See, D. KEHL – P. GUO – S. KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing, Responsive Communities Initiative*, 17.

⁶¹¹ See, L. D’AGOSTINO, *Gli algoritmi predittivi per la commisurazione della pena. A proposito dell’esperienza statunitense nel c.d. evidence-based sentencing*, 360.

⁶¹² D. KEHL – P. GUO – S. KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing, Responsive Communities Initiative*, 13; B. HARTCOURT, *Assessment program*, 2005, 32.

that there is a direct link of proportionality between the social dangerousness and the re-education of the offender, such that individuals with a low risk of reoffending are those who are more easily considered as good 'candidates'⁶¹³ for the rehabilitation phase.

On the contrary, on the contrary, the acknowledged aptitude to commit offences of some individuals would make the need for re-socialisation retreat in the face of the need for their long-term (or even permanent) incapacitation in order to protect public safety. It is precisely this thesis that recalls the positivist idea of the special-preventive efficacy of punishment, which nevertheless leaves the field open to various doubts and questions.

In this respect, in fact, there is no well-founded scientific evidence that can confirm the positive effects of long imprisonment on the individual's likelihood of reoffending; what is relevant from a more practical point of view is that it is not necessarily the case that as the prison sentence increases, the likelihood that the offender will reoffend decreases. Indeed, once the results of the algorithmic risk assessment have been received, it would be more likely that the final decision would ultimately depend on the judge's personal convictions as to the purpose of the sentence. Indeed, it would be more appropriate for such tools to be used to support sentencing in a critical and constructive manner. Indeed, the judge should consider the purpose of the criminal sanction and then assess the actual usefulness of such instruments and the real impact of the risk score on the quality and quantity of the sentence to be imposed in the individual case.

5.1 The intersection of two provisions at the stage of assessing the penalty treatment

It should be pointed out that in the Italian legal system there are two provisions dealing with and dealing with the delicate phase of the choice of penalty: Article 132 and Article 133 of the Criminal Code.

In fact, while on the one hand the first provision imposes on the judge the power-duty to determine the penalty discretely, within the limits set by the law, while obliging him to indicate the reasons justifying the use of such a discretionary power, on the other hand, Article 133 of the Criminal Code provides that, in exercising this discretionary power⁶¹⁴, the judge is bound

⁶¹³ L. D'AGOSTINO, *Gli algoritmi predittivi per la commisurazione della pena. A proposito dell'esperienza statunitense nel c.d. evidence-based sentencing*, 360.

⁶¹⁴ "Attribution to the judge of the discretionary power to determine the penalty, legal regulation of the criteria for the exercise of said power and the obligation to state reasons to guarantee the correct application of such criteria, represent, therefore, the foundations, circularly connected and logically interdependent, of the discipline in question, in the mirror of constitutional principles. To the extent to which the requirement of the legality of the penalty is intertwined with demands for areas of applicative discretion, which take account of particular and individualising aspects of the concrete case". See, D.PULITANÒ, *Diritto penale*, Turin, 528.

to follow a series of legal criteria that remain polarised around two particular figures: the seriousness of the offence and the offender's capacity to commit offences⁶¹⁵.

On closer inspection, precisely in relation to the commensuration of the penalty in the strict sense, Article 133(2) of the Criminal Code requires the formulation of a criminological prognosis concerning the assessment of the offender's capacity to commit offences. The judge, when called upon to identify the penalty to be inflicted in concrete terms, must take into account (among other factors) the future conduct of the offender. Having ideally identified the proportionate punishment for the act committed, the judge may inflict a quantum of punishment below the proportion in order to satisfy special prevention requirements. And it is precisely this finalistic orientation in the commensuration of the penalty, which has a constitutional basis in the principle of re-education laid down in Article 27(3) of the Constitution, that requires a prognostic assessment of recidivism.

In the mirror of the constitutional principles, in the discipline and on the slippery slope on which we shall move throughout the examination of this chapter, we see how the attribution to the judge of the discretionary power to determine the penalty, the legal regulation of the criteria for the exercise of that power and the obligation to state reasons to guarantee the correct application of those criteria, represent, without doubt, fundamentally connected but logically independent criteria. Therefore, to the extent to which the requirement of the legality of the penalty is intertwined with instances and spaces of applicative discretion that take into account particular and individualising aspects in the concrete case, one can see how the discretionary power of commensuration of the penalty, which represents a central moment of the entire penal system, 'marks the point of *maximum* tension with the principle of legality that inspires criminal law'⁶¹⁶.

It should be noted as of now that the subject of prognosis in the penalty system has been the subject of less in-depth study than the role of prognosis in the general theory of crime. However, it is a theme that is of fundamental importance, because it profoundly conditions the punitive system: not only in relation to the second track (today less trafficked) of security measures and the controversial concept of social dangerousness, but also in reference to the main track of punishment both in theory and in practice. And it is also for this reason that, as already mentioned, research and study on this topic has focused more on the prognoses that belong to the system of punishment.

⁶¹⁵ F. MANTOVANI, *Manuale di Diritto penale*, 773.

⁶¹⁶ F. PALAZZO, *Corso di diritto penale. Parte generale*, 582.

Starting from the assumption of the indeterminateness and incompleteness of the criteria provided by the two provisions, reflection has moved in an attempt to provide the judge with a necessary step that moves towards the use of certain instruments that can only 'help' him in the choice of the best treatment that is individualised and tailored to the individual.

In particular, when one speaks of 'individualising' the penalty treatment or the choice on punishment, it means making a judgement on the special-preventive effect of the response to the crime, in terms of containing the risk of reoffending. It is to go beyond the retrospective look at the offence, which is the point of view taken by the retributive theory of punishment. If one looks at the offender and his future life choices, the criminal offence takes on a relative meaning. The punitive response to that fact must necessarily come to terms with the objective of preventing the commission of further offences. If one replaces the photograph of the past (retribution) with a journey into the future (re-education), then prognosis takes on an irreplaceable role in the punitive system. And this is precisely the fundamental building block for realising the constitutional purpose of punishment.

What is certainly noticeable is that the gradual expansion of mechanisms that modify the substance of the sentence imposed by the legislature and the one concretely inflicted by the judge has shattered the ideal intangibility of the offence-penalty sequence. The penalty imposed increasingly resembles a project that takes shape and changes according to the characteristics and behavioural evolution of the offender. An evolution that can only be assessed in a prognostic dimension. It is no longer the idea of a just penalty because it is simply proportionate to the offence that governs the punitive system, but rather a "mobile" penalty, "in the making" which, starting from the measure established by the judge of cognition, adapts to a given individual, to his or her unique qualities and characteristics, as well as to his or her specific course of treatment. It goes without saying that this becoming of the sentence requires prognosis, i.e. (inevitably hypothetical) judgements on the future behaviour of the offender. Judgments on the basis of which it is possible to modify the content of the response to the offence, and then progressively open to the offender those spaces of freedom that accompany him towards the minimum objective of re-education: a life conduct that is respectful of penal precepts. It is precisely considered that the idea of this project-penalty must be individualised in the measure and may, if necessary, be conditionally suspended or modified in content before and during execution. In particular, a penalty that therefore modifies and adapts through a choice of readjustment on the individual. This is considered possible only if one goes through

a single type of judgement: the prognostic judgement on recidivism or on the special-preventive effects of the main or substitute punitive response⁶¹⁷.

In fact, in order to better understand the teleological and methodological profiles of the delicate commensurative phase, it appears preliminarily all the more opportune to consider the specific meaning assumed by the notion of judicial discretion in criminal matters in order to reach the conclusions of final proposals.

6 *Criminal discretion in the Italian legal system: the difficult framing*

The attention that is paid to the concept of penal discretion already overcomes all the conceptions linked to a retrograde idea connected to the 'sovereignty of the judge' or understood as a 'power of indulgence' understood as a sort of subjective equity, inspired by considerations of expediency⁶¹⁸. In fact, in order to extrapolate the concept of discretion, as it is understood today, one must start precisely from the orientations that intend this notion as a "power of choice that is bound or free as the case may be or that generically refer the same to the indeterminateness of the linguistic sign of the case in question, so that discretion takes on the characteristics of typicality *per relationem*, according to a process of heterointegration that revolves around the incompleteness of the normative model"⁶¹⁹.

The true essence of criminal discretion has been discussed, which cannot be obtained a priori from other sectors of the legal system, but should be derived from the only datum to which it can be finalistically linked, i.e. the provision laid down in Article 133 of the Criminal Code, since only through the rigorous application of this article can a systematic and unitary evaluation of the concept be achieved.

Starting precisely from such legal premises, it is noted how the essence of the criminalistic notion of discretion is identified in the "recognition by the legislator of the impossibility of foreseeing in advance in the form of law, in a general and abstract manner, the infinite value structures of the fact that translate into the greater or lesser reprehensibility of the agent, i.e. the significance of value or *disvalue*"⁶²⁰. And it is precisely by following this perspective that discretion is distinguished from the level of interpretation of indeterminate, elastic or vague

⁶¹⁷ Moreover, to this must be added that even the longed-for and never realised extension of the catalogue of principal punitive responses is confronted, already at the time of formulation of the edictal framework, with special prevention purposes whose verification, both in the abstract and in concrete terms, can only be carried out on the basis of prognostic judgements.

⁶¹⁸ G. DE FRANCESCO, *Diritto penale. Principi, reato, forme di manifestazione*, Turin, 2022.

⁶¹⁹ F. BRICOLA, *La discrezionalità nel diritto penale, vol. I, Nozione ed aspetti costituzionali*, 57.

⁶²⁰ F. BRICOLA *La discrezionalità nel diritto penale*, 58 ss.

concepts and is identified in a judicial source of criminal law, which is creative in terms of value meanings: indeed, the distinguishing mark of the concept under consideration is identified in the defect of a given abstract signifier and in the explicit legislative reference to the concrete case, to the particularities of the fact and of the perpetrator, given that certain elements do not possess an unambiguous and abstract meaning in a positive or negative sense⁶²¹.

Thus, on the notion of discretion as a normative paradigm expressed in discretionary form, the teleological conception is founded, according to which in the presence of such a datum, the judge must first identify the purpose of the institution and therefore seek the value meaning of the concrete case, i.e. the value congruent with the purpose for which the discretionary power is conferred, in order to avoid total judicial arbitrariness⁶²².

Therefore, if discretion therefore expresses the need for the judgement of the concrete case in terms of equitable justice, the control on the conformity of the purpose, on the observance of possible and possible extrinsic limits, and on the logical rigor of the process of extracting the meaning of the concrete case, to be exercised through the motivation, becomes fundamental⁶²³.

Indeed, the need for the evaluative operation to take place within and in compliance with the circle of values, intrinsic to the legal nature and purpose of the treatment to be adopted, in accordance with the provisions of Articles 25 and 27 of the Constitution, takes on a significance as a guarantee that welds criminal discretion to the principle of legality and the personalistic profile of the criminal offence.

Indeed, the nexus between individualisation, the sanctioning consequences of the crime of punishment in relation to the specific needs of the concrete case and the constitutional principles of equality, personality of responsibility and re-educative finalism, lies at the basis of the tendential constitutional illegitimacy of the so-called *fixed penalties*⁶²⁴.

The problem of the commensuration of punishment in concrete terms revolves precisely around the guarantee requirement of legality, the principle of bound discretion and that of so-called *free discretion*⁶²⁵. On closer inspection, the fact that it is the binding nature of penal discretion that characterises and distinguishes it from administrative discretion, which is

⁶²¹ *Ibidem*, 193 ss.

⁶²² *Ibidem*, 144.

⁶²³ *Ibidem*, 208 ss.

⁶²⁴ C. cost., sent. 14.4.1980, n. 50, in *Riv. it. dir. proc. pen.*, 1981, 725, con nota di C. E. Paliero, Pene fisse e costituzione: argomenti vecchi e nuovi, 1891, 725; See also F. MAZZACUVA, *Le pene nascoste. Tipografia delle sanzioni punitive e modulazione dello statuto garantistico*, Turin, 2017.

⁶²⁵ F. MANTOVANI, *Manuale di Diritto penale*, 773.

centred on assessments of expediency and convenience, emerges clearly both from the preparatory work and from the subsequent positions taken by the Constitutional Court with regard to the coessentiality that must be found in criminal matters between penal discretion and the legality of punishment⁶²⁶.

Indeed, it is observed that as a tendential balance between the need for certainty and predetermination of the penalty expressed by the rigidity of the punitive response (the fixed penalty) and the indeterminateness of the sanction, entrusted to criteria of opportunity, equity and utility assessed from time to time by the judge, the principle of constrained discretion aims to reconcile the need for certainty and legality with the simultaneous need for proportion/individualisation of the penalty, in accordance with the principle of the personality of criminal responsibility and the re-educational purpose of the criminal sanction⁶²⁷.

6.1 *Sentence commensuration and criticised discretion*

As has already been said, the judge's discretionary power in the phase of commensuration of punishment represents a 'broad and boundless theme, filtering practically through the prism of punishment the whole crime and the whole person of its author'⁶²⁸. Indeed, if the

⁶²⁶ After observing that "the principle of the legality of punishment does not tend to make it foreseeable what penalty is incurred for each offence, nor does it imply that the law must rigidly determine the penalty to be inflicted concretely" [C. cost., sent. 12.3.1962, no. 15, in Giust. cost., 1962, no. 15, in Giust. cost., 1962, 161], and that "in criminal law the system, with the exception of exceptional cases of fixed penalties, cannot achieve an adequate correspondence of the sanction to the illicit fact, except by means of the concrete assessment of the individual case, and with that determination of the penalty that, from time to time, with regulated discretion is made by the judge" [C. cost., sentence no. 25, 9.3.1967, in Giur. cost., 1967, 195], the Constitutional Court states that "the principle of legality cannot disregard the individualisation (of the penalty), that is, its adaptation to the individual case" [Const.] In this way it is noted that "the adjustment of punitive responses to concrete cases - in terms of equality and/or differentiation of treatment - contributes, on the one hand, to making criminal responsibility as "personal" as possible, in the perspective marked out by Article 27, paragraph 1, of the Constitution; and at the same time it is an instrument for determining the penalty as "targeted" as possible, in the perspective of Article 27, paragraph 3, of the Constitution". [C. cost., sent. 14.4.1980, n. 50, in *Riv. it. dir. proc. pen.*, 1981, 725 ss.; in the same direction, see also C. cost., sent. 24.6.1992, n. 299, in *Riv. it. dir. proc. pen.*, 1992, 1468, con nota di GROSSO; C. cost., sent. 23.5.1991, n. 285, in *Cass. pen.*, 1992, 22; C. cost., sent. 23.4.1991, n. 203, in *Cass. pen.*, 1991, I, 1935]. In the same vein, the jurisprudence of legitimacy has specified that, for the purposes of sanctioning treatment, the provisions of Articles 132 and 133 of the Criminal Code, in the impossibility of cataloguing the various elements of value, undeniably provide for 'discretionary spaces', even if these are of a binding nature: that is, they do not focus - as in administrative activity - on reasons of expediency, there being not only well-defined limits, but legal criteria guiding the judge's power; [Cass., Sez. I, 21.8.1990, Minic Bozidar, in *Riv. pen.*, 1991, 666].

⁶²⁷ The constitutional guarantee of the criminal court's discretionary power and the teleological functional connection with the principle of legality also form the basis of the affirmation of the principle according to which sentencing provisions are constitutionally legitimate, provided that the breadth of the sentencing range does not exceed the margin of flexibility required to allow the individualisation of the penalty in accordance with the criteria of Article 133 of the criminal code. 133 of the Criminal Code: in fact, when the edictal framework is excessively wide "the legislative predetermination of the measure of the penalty becomes merely apparent and the power conferred on the judge is transformed from a discretionary power into an arbitrary power" [so see C. cost., judgment 24.6.1992, no. 299, in *Riv. it. dir. proc. pen.*, 1992, 1468.

⁶²⁸ Così, sul punto, G. VASSALLI, *Il potere discrezionale del giudice nella commisurazione della pena*, 1316.

commensuration of the penalty per se should really represent the great test of *theories on punishment*⁶²⁹, as already mentioned, the starting point⁶³⁰ is precisely the two general clauses provided for in the provisions of Articles 132 and 133 of the criminal code⁶³¹.

In particular, these are two provisions that are closely related to each other, but which nevertheless appear to be logically complementary and functionally interdependent; while on the one hand, the provision ex art. 132 of the Penal Code provides for the duty to state reasons as a guarantee of the duty and congruity with respect to the purpose of the discretionary assessment, the second, on the other hand, represents the only instrument for the choice and graduation of the penalty⁶³². Indeed, as has already been reiterated, the expansive and systematic force possessed by these two provisions is such as to be able to transcend the theme of the application of the penalty for which the provision was conceived; from this very point of view, Article 133 of the Criminal Code takes on in the system the significance of a veritable paradigm of penal discretion, of a fulcrum around which the present provisions of a discretionary nature revolve.

In particular, starting precisely from the assumption that the discretionary power is articulated in all those possibilities of choice that the law allows the criminal judge in relation to punishment, a distinction is made between commensuration of the penalty in the strict sense, which concerns the quantification of the basic penalty within the limits of the sentence and the choice of the species of penalty in the case of alternative sentences, and commensuration in the broad sense⁶³³, which includes all the further and different hypotheses of discretion relating, however, to the sphere of punishment.

⁶²⁹ C. PEDRAZZI, *Introduzione al diritto penale*, Milan, 603.

⁶³⁰ L. MONACO- C. E. PALIERO, *Variazioni in tema di crisi della sanzione*: la diaspora del sistema commisurativo, 1994, 422.

⁶³¹ In particular, according to Bricola, the substantive and procedural regulation of the institution is found within these two provisions. See on this point, F. BRICOLA, *La discrezionalità nel diritto penale*, 73 ss.

⁶³² *Ibidem*, 105 e 144 ss.

⁶³³ The latter field presents a casuistry of considerable extension and articulation, which can, in principle, be divided into four areas of the exercise of discretion, the source of which is provided by substantive, procedural and prison legislation. Firstly, the field of hypotheses characterised by the assessment of the merits or otherwise of the application of a given institution or a given measure, in terms of negation/affirmation (*in bonam partem*: the grant or otherwise of suspended sentences, judicial pardon, payment in instalments of fines, etc.; *in malam partem*: the optional revocation of suspended sentences, the application of recidivism, the assessment of social dangerousness, the judgment on the declaration of habituality, etc.). Secondly, the cases in which the judge is called upon to carry out a quantification other than that relating to the basic penalty (determination of the quantum of increase or decrease in the basic penalty in the presence of a circumstance, of the quantum of increase in the penalty quantified for the most serious breach in the case of a continuing offence or formal concurrence of offences, of the quantum of decrease in the penalty in the context of plea bargaining, calculation of the sentence to be served in the event of revocation of probation to community service, etc.). Thirdly, the judicial search for elements of value (or disvalue: so-called undefined aggravating circumstances) not specified in the legislation, capable of influencing the penalty treatment (e.g. the identification of general extenuating circumstances, the cases referred to in Articles 114, 609-bis (3), 648(2) of the Criminal Code, 73(5) of the Narcotics Code, etc.). Lastly,

When the judge finds himself deciding on the type and measure of the penalty to be imposed he must obviously take into account, among the parameters, the future conduct of the offender.

Indeed, on closer inspection, it is to be noted that the very provision that traces the theme of the judge's discretion should be able to provide in its entirety criteria that are applicable to the judge in assessing his discretion.

In fact, however, as already mentioned, there are many cases in which the prognostic evaluations entrusted to the judging body in the Italian penal system are not anchored in firm and objective criteria to which it can refer⁶³⁴.

In this regard, reference is made to the theoretical construction of prognoses and the structural margin of uncertainty that surrounds them, but, at the same time, also to the input and scrutiny of scientific knowledge (statistical, criminological, etc.), which is necessary to carry it out. Therefore, on this point, there are still some questions that remain open and which concern: the temporal vagueness of the prognosis (the possibility that the subject may commit other offences in the future), the vagueness of their subject matter, deriving either from their own or from the lack of a more detailed and circumscribed catalogue.

Precisely in order to address these gaps in the system, it is believed that some jurisdictions have chosen to equip themselves with systems (in some cases with I.A. features) that are capable of filling them.

Among the criteria for the commensuration of punishment, provided for in Article 133(2) of the Criminal Code, is the capacity to commit offences. However, the greatest problems are to be found in the fact that the legislator has not provided a definition of this concept but has identified numerous factual criteria on the basis of which the judge must reconstruct the offender's capacity to commit offences: the offender's personality, conduct prior to,

hypotheses in which the judge is called upon to choose one of two or more alternative solutions, which directly affect the punitive treatment (as is the case, for example, of the judgment of comparison in heterogeneous concurrences of circumstances) or which are connected to it in a potential way (as is the case when choosing the precautionary measure), can be brought within the area of commensuration in a broad sense. While recognising that, depending on the discretionary structure of the institution under consideration, the specific function and rationale underlying it are of relevance, the dominant opinion is that, regardless of the express reference, it is Article 133 of the Criminal Code that acts as a general and paradigmatic model of discretion in all hypotheses of commensuration in the broad sense. In doctrine, however, it has been observed that Article 133 of the Criminal Code is not capable of providing a unitary discipline of the judge's discretionary power, since commensuration in the broad sense must in these cases necessarily appeal to criteria further and different from those indicated by the aforesaid provision. On this point, F. PALAZZO, *Corso di diritto penale. Parte generale*, 592; and also, T. DELOGU, *Potere discrezionale del giudice e certezza del diritto*, 397.

⁶³⁴ In this regard, one thinks not only of the theoretical construction of prognoses and the structural margin of uncertainty that surrounds them, but also of the input and scrutiny of scientific knowledge (statistical, criminological, psychological, etc.) required to make them. In this regard, it will suffice to mention, merely by way of example, some of the questions that remain open: the temporal vagueness of the prognosis (will he commit or refrain from committing further offences in the future); the vagueness of their object, resulting from the failure to identify a circumscribed catalogue.

contemporaneous with, and subsequent to the offence), and individual, family and social living conditions. While it is true that there is a divergent view of the concept of capacity to commit offences, it is also true that the preferable interpretation would seem to be that which identifies this concept in the capacity to commit future offences. This is because it remains anchored to the constitutional dictate; indeed, if punishment is to respond to re-educative needs - pursuant to Article 27(3) of the Constitution. - it is then necessary to take into consideration the specific characteristics of the individual.

Indeed, it is also necessary to consider that the purpose of special prevention encounters a limit at the commensurative stage: in fact, it cannot exceed the guarantee limit which, pursuant to Article 27, paragraph 1 of the Constitution, is marked by the proportion to the culpability for the fact⁶³⁵.

.In other words, the judge's task is multifaceted in several stages: indeed, after determining the sentence proportionate to the gravity of the concrete fact, pursuant to the provision under Article 133(1) of the Criminal Code, the judge eventually identifies the amount of the lesser sentence useful to meet the subject's re-education/resocialisation needs, pursuant to Article 133(2). In other words, the criminal capacity of the subject may possibly be relevant only *in bonam partem*: in the dosimetry of the penalty, that is, the judge must consider the possibility, in more or less maximum terms, that the subject may commit a crime again and, consequently, reduce the penal sanction commensurate with the seriousness of the crime committed. However, the opposite does not apply: that is, an increase in the concrete penalty by reason of the greater 're-educational needs' of the offender.

6.2 *The 'capacity to commit offences' as an assessment that forces one to look into the future*

The debate on the nature and characteristics of the assessment of criminal capacity has focused on the attempt to define the type of assessment with which the judge has to deal on a daily basis.

There are undoubtedly elements that point towards a prognostic reconstruction of the criminal prognosis of the subject⁶³⁶. In the first place, in fact, it is certainly appropriate to look

⁶³⁵ L. EUSEBI, *Tra crisi dell'esecuzione penale e prospettive di riforma del sistema sanzionatorio: il ruolo del servizio sociale*, in *Riv. it. dir. pr. pen.*, 1993, 498 ss.; G. FIANDACA, *Commento all'art. 27, comma 3, in Commentario alla Costituzione. Rapporti civili*, in G. Branca - A. Pizzorusso (eds), Bononia, 1991, 327 s.; L. MONACO, *Prospettive dell'idea dello "scopo" nella teoria della pena*, 108; L. MONACO - PALIERO, *Variazioni in tema di "crisi della sanzione"*, 434 ss.

⁶³⁶ This expression seems more correct when one considers that the capacity to commit crimes concerns all offences and not only crimes F. ANTOLISEI, *Manuale di diritto penale. Parte generale*, Milan, 1955, 455.

at the historical origin of the rule since the projection towards the future of this concept would appear more consistent with the ideas of the positive school, to which the discipline of criminal capacity seems to be inspired⁶³⁷. Indeed, another element in favour of this interpretation can also be seen in the concepts employed by the legislature⁶³⁸.

On closer inspection, a heated doctrinal debate has been built and broadened over time around the concept of the capacity to commit offences; a debate which, in reality, responds to and reflects the different ideologies that permeate the penalty system (and more generally), the entire penal system; in fact, a retributive vision centred on the fact committed is contrasted by a projection on the future behaviour of the subject in terms of preventing recidivism. In fact, even if, on the one hand, a reconstruction in a preventive perspective of the subject's capacity to commit offences in the light of the constitutional dictate appears preferable, it is undoubtedly impossible to conceal how the lack of an express finalistic indication on the part of the legislature can frustrate these reconstructions.

In fact, the so-called "emptiness of purpose", within the meaning of Article 133 of the Criminal Code, and also the ambiguity of the various factual indices described by the rules, ultimately undermine the binding nature of the commensurate criteria.

Indeed, in order for there to be an effective 'rational' exercise of the discretion entrusted to the judge (and not only in the commensuration of the penalty) it is also necessary for the legislature to identify precise criteria for the exercise of the same discretion to which it refers⁶³⁹.

Indeed, and on this point an independent reflection is needed, it is precisely the criterion of the 'seriousness of the offence' or the assessment of the 'capacity to commit offences' that are types of assessment that fail to guide the judge precisely because there is no legislative choice as to what the aim is actually pursued by the penalty. Indeed, as has been carefully observed, the discretionary power of commensuration of the penalty entrusted to the judge is 'lacking a compass. Indeed, markedly without a compass, [...] in the sense that the judge is not even provided with the teleological parameter of commensuration'⁶⁴⁰. To demonstrate this thesis, it would seem sufficient, indeed, to cite the practice with regard to the motivation: in fact, the reasoning that the judge follows in choosing and, consequently, in commensurating the

⁶³⁷ In this regard, in fact, as confirmed by the Ministerial Report to the final draft of the Criminal Code: 'the dangerousness of the offender in so far as, in the application of punishment, it coincides with the capacity to commit offences, i.e. with the aptitude of the individual to violate criminal law'. Thus, as also reported by A. MALINVERNI, *Capacità a delinquere*, in *Enc. Dir.*, Vol. VI, Milan, 1960, 125.

⁶³⁸ D. PULITANÒ, *Diritto penale*, 474.

⁶³⁹ See for all F. BRICOLA, *La discrezionalità nel diritto penale*, 80 ss.; E. DOLCINI, *La commisurazione della pena*, 177 ss.; L. MONACO, *Prospettive dell'idea dello "scopo" nella teoria della pena*, 208 ss. and 267 ss.

⁶⁴⁰ T. PADOVANI, *La disintegrazione attuale del sistema sanzionatorio e le prospettive di riforma: il problema della comminatoria edittale*, 427.

concrete penalty is completely inaccessible. In fact, as already highlighted above, the so-called "lazy" or "stylistic" wording (e.g. "it is considered fair") certainly does not make it possible to retrace the argumentative procedure followed by the judge in order to arrive at the choice and quantification of the concrete penalty, which would appear to depend on and appear to be left to the sensitivity and intuition of the individual judge⁶⁴¹.

6.3 *The problem of prognostic evaluations*

The subject of prognostic judgments constitutes a problematic knot that touches the delicate and changing balance between legislative and judicial power in the determination of punishment. In fact, if prognoses in the penalty system immediately evoke the judge's discretion and the individualisation of the penalty according to the characteristics of the concrete case, it is nevertheless the legislator who plays a decisive role. And this is not only because it is up to the legislature to establish the purposes, application spaces and operating rules of prognostic judgments in the penalty system, but also because the more or less recent history of criminal law is studded with legislative interventions aimed at neutralising the judge's discretionary power in the formulation of prognoses through the provision of rigid presumptions. What of prognostic judgements remains a perennially contested territory between legislative constraints and the judge's free conviction. The first sometimes translates into legal evidence on the future behaviour of the offender, as happened, for example, in the case of presumptions of social dangerousness or, more recently, in relation to the mandatory application of recidivism in the case of the commission of certain types of offence. The second, i.e. free conviction, represents in some ways, the antidote to legislative automatisms, but in turn risks being inevitably vitiated by the judge's intuition, emotionalism and arbitrary generalisations. It is precisely this law-judge dialectic that lies at the heart of prognostic judgments. One need only glance at the constitutional jurisprudence to realise which and how many prognostic judgments have been subtracted from the judge's discretion, i.e. from his free conviction, to be entrusted to actual legal evidence. In these cases, the legislator completely subtracts from the judge the prognosis and, with it, the evaluation of the evidence that allows one to make predictions on the future behaviour of the offender⁶⁴². That the instrument for realising preventive purposes should be taken away from the power of the judge, i.e. the one

⁶⁴¹ E. AMODIO, *Motivazione della sentenza penale*, in *Enc. dir.*, XXVII, Milan, 1977, 229 s.; F. BRICOLA, *La discrezionalità nel diritto penale*, 109 ss.; E. DOLCINI, *La commisurazione della pena*, 59 ss.; S. LARIZZA, *La commisurazione della pena: rassegna di dottrina e giurisprudenza*, in *Riv. it. dir. proc. pen.*, 1982, 604.

⁶⁴² P. FERRUA, *Un giardino proibito per il legislatore: la valutazione delle prove*, in *Quest. Giustizia*, 1998, 587 ss.

who should be called upon to execute the programme for the purpose of punishment, seems to be a real contradiction in terms. It is not, however, uncommon for legislators to distrust judicial discretion, since the objective of rendering justice in the concrete case does not always (or perhaps almost never) coincide with the pursuit of general preventive purposes. Making a prognosis as to the future behaviour of the offender, in order to best realise the purposes of special prevention, means possibly renouncing the infliction of punishment or changing the content of the sanction abstractly threatened by the legislature. And this brings with it the suspicion that punitive power has been abdicated, to the detriment of the certainty and effectiveness of punishment. The legislative presumptions on the subject of prognosis limit these (apparent?) risks of loss of general-preventive efficacy of the threat of punishment. In order to try to understand the role of legislative prognoses in the penalty system, the analysis is intended to move from a privileged observation point: that of the constitutional jurisprudence on legislative automatism in relation to the effectiveness of punitive responses in the prevention of recidivism⁶⁴³. In conclusion, one observes how the effectiveness of the entire penalty system depends, at least in part, on the delicate issues connected with prognostic judgements⁶⁴⁴. It is considered useful and necessary to reflect on the subject because it is also believed that if prognostic instruments are not refined, the individualisation of punishment remains little more than a political-criminal wish and the catalogue of punishments becomes an *instrumentarium* whose effects and special-preventive efficacy are unknown.

6.4 *The answer to a question: why prognosis is considered so important in the choice of sanction treatment*

It has been observed that 'the satisfaction of the needs of social defence, which have not been met by the ever more exorbitant security measures, '[has] been sought within the punishment: the real double track, that is, the one that counts, because it is on this that the legislator intends to play the wager of the effectiveness of the penalty system, is no longer the one devised by criminological positivism, but the one that translates into the provision of differentiated courses of punishment, both in the commensuration criteria and in the developments in the executive phase⁶⁴⁵. This genetic mutation of the double binary and the

⁶⁴³ In fact, the Constitutional Court's interventions make it possible, on the one hand, to identify the balance between legislative and judicial power in the formulation of prognostication and, on the other hand, offer a fundamental theoretical framework for checking the empirical, factual tightness of prognostic judgments and, therefore, ultimately their reasonableness.

⁶⁴⁴ See C.E. PALIERO, *Il principio di effettività nel diritto penale*, Naples, 2011.

⁶⁴⁵ M. PELISSERO, *Pericolosità sociale e doppio binario*, 69.

consequent multiplication of differentiated punishment paths (intra- and extra-prison) claims an extremely relevant role for prognosis: the historically most decisive one because it no longer involves only security measures, but first and foremost punishment. And it is for these reasons that we have chosen to focus the study of prognosis on the side of punishment.

Indeed, although it represents the needle of the scales both in the phase of commensuration of punishment and for the application of alternative sanctions to prison (such as probation, probation, alternative sanctions and alternative measures), the in-depth study devoted to prognosis in relation to punishment is very limited and the situation is even more bleak if one shifts attention to practice.

In fact, to formulate hypotheses on future events, the legislator often relies on rigid and automatic presumptions, which impose forms of neutralisation and repression. When instead relying on the judge's discretion, jurisprudence takes refuge, in the vast majority of cases, in intuition and common sense. This means, in fact, systematically evading the problem of prognoses, which remain vague normative clauses to be filled in discretionally (or worse, arbitrarily) with the personal ideas of the person making the prognosis and who does so, moreover, through very little or almost non-existent motivation⁶⁴⁶.

It is precisely the duty to state reasons that should 'compel the judge to decide rationally'⁶⁴⁷. It is, however, a constraint that only works if the wording of the law succeeds, on the one hand, in anchoring the prediction to scientific criteria and, on the other hand, in fostering a jurisprudential practice that does not feel entitled to take refuge in the irrationality of intuition. In case law one mostly finds apodictic and peremptory statements, which only reiterate, in their obvious fragility, the already vague legislative formulations. If the formulation of prognostic judgments is purely formal, they risk becoming an 'empty shell' that conceals the implementation of a criminal policy programme of the individual magistrate, outside of any (and mostly laconic) legislative indication. This is what happens whenever the judge justifies on the basis of so-called style clauses the choice and quantification of the concrete sentence or the granting of the suspended sentence. Relying on emotional intuition exponentially increases the risk of uncertainty and precariousness that is already inherent in prognostic judgments. As has been observed, 'intuition proceeds by leaps and bounds to self-evident truths. Motivation proceeds by degrees, through arguments, up to truths that are never self-evident and that,

⁶⁴⁶ Le formule vaghe utilizzate dal legislatore nel descrivere i giudizi prognostici all'interno dell'intero ordinamento vengono concretizzate dal giudice mediante «operazioni di fantasia creativa», *cfr.* M. TARUFFO, *Sui confini. Scritti sulla giustizia civile*, Bononia, 2001, 332 ss.

⁶⁴⁷ F.M. IACOVIELLO, *La Cassazione penale. Fatto, diritto e motivazione*, Milan, 2013, 298.

indeed, never lose the character of problematicity. The *intimate conviction* is a psychological event, motivation a logical phenomenon"⁶⁴⁸. Through the irrationalistic version of self-belief, i.e. the one based on pure emotional intuition, one runs the risk of producing negative consequences on the prevention of reoffending and social defence, because negative prognoses can be reached in respect of subjects who will then commit further offences, or, and in the opposite direction and with far more serious repercussions, to impose (avoidable) limitations of personal liberty on persons who, on closer scrutiny (in terms of cognitive background and scientific instruments of assessment), should have been the recipients of a favourable prognosis. The more the effect of the prognostic judgment affects a person's fundamental rights, as in the cases envisaged by the penal system, the more the prognosis, though inevitably lacking absolute certainty, must be sustainable and justifiable: recourse to personal intuition seems to disregard this banal guarantee. Not only that: recourse to intuition seems all the more convenient and probable the less information is available to the judge. As has been observed, 'paradoxically, it is easier to construct a coherent story when little is known and there are fewer pieces to fit into the puzzle. Our comforting belief that the world is endowed with meaning rests on a secure foundation: our almost unlimited capacity to ignore our own ignorance'⁶⁴⁹. This seems to be the trap into which the judge falls when he formulates the prognosis on the basis of uninformed intuition. In reflecting the exiguous and sporadic relations between empirical sciences and criminal law⁶⁵⁰, the scant attention devoted to the role of prognostic judgements in the penalty system leaves open numerous problematic questions: on the method for making them; on the choice of factors to be taken into account; on their motivation and verifiability; and on the general principles governing them (we intend here to refer to predictive factors and the attempt to explain criminal behaviour on the basis of established experience). In this respect, one thinks not only of the theoretical construction of prognoses and the structural margin of uncertainty surrounding them, but also of the input and scrutiny of scientific knowledge (statistical, criminological, psychological, etc.) required to make them. In this regard, it will suffice to mention, by way of example only, a few questions that remain open the indefiniteness of the prognosis in time (will he commit or refrain from committing further offences in the future); the indefiniteness of their subject matter, resulting from the failure to identify a circumscribed catalogue of offences to be committed the indefiniteness of the rules of judgement resulting, on the one hand, from the absence of balancing rules in the

⁶⁴⁸ *Ibidem*, 300.

⁶⁴⁹ See D. KAHNEMAN, *Pensieri lenti e veloci*, Milan, 2012, 221.

⁶⁵⁰ G. ZARA, *Valutare il rischio in ambito criminologico. Procedure e strumenti per l'assessment psicologico*, 17.

(very frequent) case that the indicators to formulate the prognosis are opposite, and on the other hand, from uncertainties as to the standard of assessment to be used (beyond reasonable doubt or preponderance of the evidence), as well as on the management of doubtful cases, which are indeed largely prevalent (whether the principle of *favor libertatis* should apply). In addition, it should be added that it is the same procedural mechanisms that play a leading role with regard to the (im)possibility of making projected findings on the future conduct of the accused/convicted person (think of the current prohibition of criminological expertise by the judge of cognition and, to the contrary, the proposals *de iure condendo* to introduce a biphasic procedural model to separate the assessment of liability from the prognosis on the effects of the punitive response).

As already observed more than thirty years ago in words that still seem relevant today, 'the problem is not how much empirical knowledge is today per se available for the knowledge of the author, but rather how much empirical knowledge is actually usable within normative structures that on the one hand have to reckon with requirements relating to practicability the time and cost of the instruments (including criminological ones) they intend to make use of, and on the other hand are obliged to respect ideological choices and principles that do not always allow all that the empirical social sciences can actually offer to be used in the process'⁶⁵¹. From a general point of view, there is certainly a major cognitive deficit that depends, on the one hand, on the lack of empirical research carried out in Italy on recidivism and, on the other hand, on the absence of procedural tools to make sufficient information flow into the trial to individualise the punitive response according to the offender's characteristics, needs and, ultimately, future behavioural developments.

The list of problematic knots to be unravelled is certainly longer; for this reason, it will suffice here to observe that the effectiveness of the entire penalty system depends, at least in part, on the solution of the delicate questions related to prognostic judgements⁶⁵². If prognostic instruments are not refined, the individualisation of punishment remains little more than a criminal-political wish and the catalogue of penalties becomes an *instrumentarium* whose effects and special-preventive efficacy are unknown.

⁶⁵¹ L. MONACO, *Prospettive dell'idea dello "scopo" nella teoria della pena*, 181.

⁶⁵² C.E. PALIERO, *Il principio di effettività nel diritto penale*.

7 *Limits and differences in algorithmic evaluation in the Italian penal system*

It cannot but be noted that in the Italian penal system, the commensuration of the penalty by the judge, as already mentioned, is an activity that is delegated to the judge's discretionary assessment. Indeed, the provision under Article 133 of the Criminal Code provides that the judge must take into account both the seriousness of the offence, as provided for in paragraph 1, and the offender's capacity to commit offences, as provided for in paragraph 2⁶⁵³.

It is precisely this latter assessment that is inferred from the offender's motives to commit offences and his character, his previous criminal record and previous conduct, his conduct contemporaneous with or subsequent to the offence and his individual, family and social living conditions.

In this regard, it may be noted that these factors overlap with those that are used to make the algorithmic assessment of the risk of reoffending in American systems⁶⁵⁴. It can be seen that, in principle, there would be a regulatory addendum to which it is possible to anchor the use of instruments in support of the judge's own decisions at the time of sentencing. However, the use of such tools would risk colliding with some basic principles of criminal procedure.

First of all, it should be noted that, unlike the American model, in the Italian system there is no biphasic distinction between the pronouncement of the sentence and then the subsequent imposition of the penalty (nor is there at the same time the intermediate phase involving the investigation of the offender's character). This is certainly one of the main differences that already place the two systems on two different levels.

Indeed, even if this particular *modus procedendi* existed or were to be established, the principle of the formation of evidence in the adversarial process of the parties, pursuant to Article 111, paragraph 4 of the Constitution, which should reasonably prevent the judge from autonomously acquiring useful elements for the purposes of commensuration of the sentence or assessing elements other than those subject to adversarial process between the parties.

On closer inspection, it is precisely the activation of investigative powers *ex officio* or even at the request of the parties that would also be precluded by the provisions of Article 220,

⁶⁵³ According to the most authoritative doctrine, the two paragraphs that make up Article 133 of the criminal code represent a compromise between the classical school and the positive school of criminal law: the former is inclined to admit the ethical-rewarding character of punishment (the offender must be punished for what he has committed), the latter to recognise its merely incapacitating function (the offender must be removed from society because he is dangerous). On the subject, A. PAGLIARO, *Commisurazione della pena e prevenzione generale*, 25; V. MILITELLO, *Prevenzione generale e commisurazione della pena*; F. BRICOLA, *La discrezionalità nel diritto penale*, 208.

⁶⁵⁴ With the exception of conduct contemporaneous with or subsequent to the crime, the others are factors that can easily be generalised statistically and made available in the form of data.

paragraph 2, of the Code of Criminal Procedure, except for what is provided for the purposes of the execution of the sentence or security measure, which provides that "expert opinions are not permitted to establish the habituality or professionalism of the offence, the tendency to commit offences, the character and personality of the defendant and in general the psychological qualities independent of pathological causes".

It follows, *de iure condito*, that it is impossible to use tools such as, for example, Compas that are based on the answers given by the defendant to the appointed consultant, unless one wishes to circumvent the prohibitions laid down by law⁶⁵⁵.

Undoubtedly, an activity of this kind, even if it were otherwise denominated, would to all intents and purposes constitute an expert report and, as such, would be unusable due to violation of the law pursuant to Article 191(1) of the Code of Criminal Procedure.

Indeed, in the very years in which American judicial practice was beginning to be confronted with the idea of *evidence-based sentencing*, the legislator in Italy of the new Vassalli Code of Criminal Procedure had already openly rejected the idea of including a preliminary investigation into the offender's capacity to commit offences⁶⁵⁶. Article 220 of the Code of Criminal Procedure indicates some of the elements that the judge should take into consideration in accordance with the provisions of the second paragraph of Article 133 of the Criminal Code, thus excluding the possibility of making technical-scientific inquiries on them. As has been observed on this point, such an approach ultimately ended up overturning the orientation in favour of expert opinion⁶⁵⁷, concerning the personality and psychological condition of the defendant, which seemed to animate the delegated law on the enactment of the new Code of

⁶⁵⁵ It could be assumed that the mere interview of the defendant is not an activity attributable to expertise. This thesis would, however, clash with the textual datum of the law, according to which an expert opinion is obtained when 'it is necessary to carry out investigations or acquire data or evaluations requiring specific technical, scientific expertise' (Article 220(1) of the Code of Criminal Procedure). In the present case, the activity entrusted to the consultant is aimed at obtaining information to acquire data and evaluations of a technical nature, and falls fully within the codified definition.

⁶⁵⁶ That the prohibition of expert opinions was also conceived with the enforcement of sentences in mind is quite clear from the words of the legislator. The wording of the prohibition does not include the phase of enforcement of the sentence, so that it would be abstractly possible for the magistrate or the supervisory court, for their respective decisions, to use actuarial risk assessment instruments without circumventing or violating the prohibition under review.

⁶⁵⁷ In the draft of the new Code of Criminal Procedure, drawn up by the Ministerial Commission chaired by Giandomenico Pisapia (the so-called Pisapia Project), the criminological expertise became a very important diagnostic tool that the judge could use whenever, in the cognitive phase, he had to formulate a judgement on the personality and social dangerousness of the offender. In fact, Article 209, paragraph 2, of the Preliminary Draft provided that 'For the purposes of the judgement on personality and dangerousness, the expert opinion may concern the defendant's personality also with regard to psychological qualities independent of pathological causes'. The subsequent art. 212 para. 2 further stated that: "Expert opinions relating to questions on personality and dangerousness shall be entrusted to specialists in criminology or to a doctor specialising in psychiatry or psychology". At that time, the usefulness of criminological expertise was also recognised by authoritative criminalists, including F. MANTOVANI, *Manuale di Diritto penale*, 670

Criminal Procedure⁶⁵⁸. There were, however, opposing positions that expressed themselves more cautiously and more favourably on the issue⁶⁵⁹.

In fact, it is peculiar to note how some authors⁶⁶⁰ who took a more rigid position on the subject immediately justified this prohibition by referring, *ex adverso*, to the provisions of Article 133 of the Criminal Code which, by leaving the commensuration of the penalty to the full discretion of the judge, would almost have been symptomatic of the codicil legislator's desire to remove the field then from the possible scientific evidence on the personality of the offender. Indeed, in order to temper the rigour of the most extremist positions, some authors⁶⁶¹ had even proposed a third possibility, which was that of the so-called biphasic investigation (which in part refers to and recalls the American model described above). On closer inspection, the distinction between the stages of the preliminary investigation would certainly bring about a great change as it would probably eliminate the danger that the criminological expertise might turn into a tool to be used against the defendant himself: in fact, if the investigation were carried out before the criminal responsibility of the offender was ascertained, the principle of the presumption of innocence, pursuant to Art. 27 of the Italian Constitution; on the contrary, in a trial characterised by a first phase reserved for ascertaining the defendant's responsibility and

⁶⁵⁸ Article 2, paragraph 1, no. 10 of Law no. 108 of 3 April 1974 provided, among the guiding principles and criteria, that the delegate should provide for "the reorganisation of the institution of expert opinion, with particular reference to medico-legal, psychiatric and criminological expertise, ensuring the highest technical and scientific competence of the experts". The express reference to criminological expertise is a clear indication of the legislator's expressed favour for the investigation of the defendant's personality. The principle was probably inserted because the legislature intended to 'break' with tradition, given that the previous Code of Criminal Procedure of 1930 laid down in Article 314(2) a prohibition very similar to that contained today in Article 200(2). The original wording of the Code of Criminal Procedure configured a process defined by the doctrine as "almost impermeable" to the contributions of the non-legal sciences. Refer to, P. MARTUCCI, *Il contributo del criminologo nel processo penale: un problema ancora aperto*, in *Diritto penale e processo*, 6, 744, who recalls that in the previous system several questions of constitutional legitimacy had arisen on the grounds of conflict with Articles 27(3) and 3 of the Constitution, on the argument that the prohibition of "personality expertise" conflicted with the principle of the re-educative purpose of punishment, in fact weakening the defendant's right to defence and resulting in different treatment of adult and juvenile defendants. However, the Constitutional Court always expressed itself recognising the conformity of the contested prohibitions with constitutional principles (Constitutional Court, 9 July 1970, no. 124, in *Riv. pen.*, 1970, II, 684; Constitutional Court, 19 December 1973, no. 179, in *Giust. pen.*, 1974, I, 72).

⁶⁵⁹ An authoritative part of the legal and medico-legal doctrine welcomed the approach of the new code, observing that the strict limit imposed on the expert opinion was intended to avoid the scientific unreliability of an examination of the defendant's personality, in view of the objective lability of the investigation and the conditioned attitude of the expert witness, and the risk of violation of the right of defence where, by limiting the psycho-physical freedom of the defendant, the guarantees and instruments typical for the acquisition of evidence could easily be circumvented. On this point, B. PANNAIN – M. ALBINO – M. PANNAIN, *La perizia sulla personalità del reo: evoluzione dottrinale e normativa. Prospettive nel c.p.p. '88'*, in *Riv. It. Med. Leg.*, 1989, 848; F. CORDERO, *Guida alla procedura penale*, Turin, 1986, 347; S. RAMAJOLI, *La prova nel processo penale*, Padua, 1995, 159.

⁶⁶⁰ V. PERCHINUNNO, "Le prove", in Pisani - Mario (eds), *Manuale di procedura penale*, 2008, Bononia, 224; P. GIANNITI, *La valutazione della prova penale*, Turin, 2005, 199.

⁶⁶¹ On this point, P. RIVELLO, *Perito e perizia*, in *Digesto delle discipline penali*, IX, 1995, 479; D. BIELLI, *Periti e consulenti tecnici nel nuovo processo penale*, in *Giustizia penale*, 1991, 65.

a second phase intended for the choice of individualised treatment (in which the criminological investigation would then be placed), there would be no risk of 'polluting' the preliminary investigation on the *an* of responsibility with evidence concerning the *quantum puniatur* instead.

7.1 *The limits posed by Article 220 of the Code of Criminal Procedure: is the principle in crisis?*

On closer inspection, as mentioned above, the provision of Article 220 of the Code of Criminal Procedure would seem to be set as a limit by the legislature to the possibility of using such tools. It should be noted that in reality the same limit imposed by the provision would not seem to exclude entirely the possibility of an algorithmic assessment of the offender's personality.

In fact, the possibility of using predictive tools would remain open, but if they were based solely on available statistical data or personal information, they would make it possible to express and calculate a hypothetical judgement on the offender's future capacity to commit offences⁶⁶².

Article 220 of the Code of Criminal Procedure is for this reason also considered the 'bulwark' in defence of the entry of algorithms into the trial, precisely because it sets the limit of the inadmissibility of expert opinions to 'establish habituality or professionalism in the offence, tendency to commit offences, the character and personality of the defendant and in general the psychic qualities independent of pathological causes. Indeed, the purpose of the expert opinion would be to draw up a profile of the offender's personality and character in order to identify the most suitable penalty or security measure for the case in question.

Another limitation lies in the fact that the prognosis of the offender's capacity to commit offences is an *intuitu personae* judgement and no statistical evaluation can support or replace such a judgement⁶⁶³.

⁶⁶² The legislature was strongly opposed to the use of such an 'automated' technique of analysis: the rationale of the prohibition can be found in the need to protect the defendant's moral freedom, since there would be a risk of yielding to the prejudices inherent in particular aspects of the defendant's character that could condition the adjudicating body. It is therefore intended to prevent the judge, in making his or her determinations, from relying essentially on the identity of the defendant drawn from the psychological reports and not on the actual facts committed.

⁶⁶³ This is also because 'the risk score would then be calculated by cross-referencing data on similar situations or events, causing the judgement on the offender's social dangerousness to fall into a labyrinth of inevitable empirical generalisations. See, A. DI PRISCO, *Elementi di criticità sulla perizia psicologica nel processo penale*, in *Ius in Itinere*, 2018.

Doubts certainly remain as to the reliability of such instruments. In fact, the verbal risk score would be calculated by cross-referencing data relating to similar situations or to analogous events, thus running the risk of dropping the judgement on the social dangerousness of the ro into a 'labyrinth of inevitable empirical generalisations'⁶⁶⁴. Nor would it seem possible to individualise the judgment through the input of detailed information on the defendant's personality, given the prohibition laid down in the second paragraph of Article 220 of the Code of Criminal Procedure.

However, it could be objected that the assessment based on statistical evidence may, depending on the reference sample, also be very reliable; or even more simply that criminal law often relies on judgements of this kind (if one thinks, for example, of the concept of statistical probability and modern theories on the legal relationship of causality)⁶⁶⁵.

However, this criticism is not entirely insuperable. In fact, the prognosis of the offender's capacity to commit a crime is precisely an *intuitu personae* judgement that is closely linked to the personality of the offender: therefore, no statistical result will ever be able to support by itself a positive or negative judgement of social dangerousness. The sample of data on delinquency history is certainly obtained by grouping offenders by different faces of age, area of residence, family situation, etc.

In such terms, in fact, an individual could be considered at risk of reoffending only by virtue of belonging to a group: this would produce the undesirable effect of 'contaminating' the judge's discretionary assessment without, however, providing useful elements to infer the offender's capacity to reoffend.

On a first reading, therefore, it would seem possible to conclude negatively on the possible application of such instruments in the Italian criminal trial. However, the limits imposed on criminological expertise would only admit an assessment based on presumptive risk indices, which are totally unsuitable for guiding judicial discretion in the commensuration of

⁶⁶⁴ L. D'AGOSTINO, *Gli algoritmi predittivi per la commisurazione della pena. A proposito dell'esperienza statunitense nel c.d. evidence-based sentencing*, 367.

⁶⁶⁵ Explanatory reasoning of mental elimination can insofar be carried out insofar as it is known beforehand that a certain action does or does not give rise to a certain event, knowledge that can be derived immediately from science; however, where the latter does not help, it must be acquired aliunde. According to the teachings of F. Stella, an antecedent can be configured as a necessary condition only if it is among those antecedents which, on the basis of a regular succession in accordance with a law endowed with scientific validity, lead to events of the kind that have occurred in concrete terms. Such general laws may be either universal laws, capable of stating that the occurrence of one event is invariably accompanied by the occurrence of another event, or statistical laws, which merely state that the occurrence of one event is accompanied by the occurrence of another event only in a certain percentage of cases, with the consequence that the latter are all the more endowed with scientific validity the more they can be applied in a sufficiently large number of cases and be confirmed by recourse to rational and verifiable methods of proof..

punishment. On closer inspection, in the Italian legal system the legislator has always maintained a certain detachment and even a reticent attitude in opening up to other branches, such as, for example, psychological and criminological science, which have always carried out studies on the personality of the subject. In conclusion, the assessment of personality, in a criminal law that over the years has struggled to see itself centred exclusively on the fact and not on the perpetrator⁶⁶⁶, has always revolved, because of the fears it engendered, around the prohibitions on the performance of criminological expertise, which are found in Article 314 paragraph 2 of the Code of Criminal Procedure and reiterated, once again, in Article 220 paragraph 2⁶⁶⁷, which certainly represent a (perhaps not insurmountable) limit for this type of assessment.

8 *Discretion and its combination of constrained and controlled in the criminal justice system*

On closer inspection, it is considered necessary to expose and start from this very starting point because it is considered necessary in order to be able to investigate within what limits and boundaries we believe can be applied, the tools of artificial intelligence in criminal law.

In fact, as we have already seen, the commensuration of punishment is, and falls within those so-called *sui generis* discretions, since it finds itself constrained and regulated, albeit within limits that are in any case considered quite broad⁶⁶⁸. The limits and boundaries within which it moves can be so-called external limits that are basically marked by the edictal framework of the individual offence; or internal limits that are marked by the provision under Article 133 of the criminal code, or even other limits of a procedural nature that are, for example, inherent in the obligation to state reasons⁶⁶⁹. These limits are usually simultaneously

⁶⁶⁶It is also interesting to note the position taken by the Constitutional Court several years ago in its judgment. After all, the Constitutional Court, in its very beautiful judgment No. 124 of 1970, saw at the basis of the prohibition of criminological expertise (also) the concern 'that the study of the personality of the defendant can only be carried out by one who is also aware of the afflictive and intimidating character of the penalty.

⁶⁶⁷ It has a close connection with Article 188 of the Code of Criminal Procedure; it is no coincidence that, to justify this limitation, Franco Cordero wrote: "there are too many *soi-disants* machinists of the soul and it is better that they do not set foot in the trial" 115. Ultimately, this is the reason why the 1988 legislature confirmed a prohibition deemed anachronistic by doctrine since the 1960s and by the Constitutional Court itself. On this subject we refer with the usual effectiveness, F. CORDERO, *Codice di procedura penale commentato*, Turin, 1990, 264; ed anche si fa riferimento a F. BRICOLA, *La discrezionalità nel diritto penale*, 116 and G. VASSALLI, *Criminologia e giustizia penale*, in G. Leone (ed), *Scritti giuridici in onore di Alfredo De Marsico*, vol. II, Milan, 1960, 581. More recently, read G. VARRASO, *La prova tecnica*, in *Trattato di procedura penale*, G. Spangher (dir), vol. II, *Prove e misure cautelari*, in A. Scalfati (ed), *Le prove*, Turin, 2009, 242-243.

⁶⁶⁸ Cass., Sez. I, 3.7.1986, Giberti, in *Riv. pen.*, 1987, 502.

⁶⁶⁹ On the nature and for an accurate reflection on these limits, see E. DOLCINI, *La commisurazione della pena*, 55; on the distinction between limits of institutional content and limits of regulatory content, T. DELOGU, *Potere discrezionale del giudice e certezza del diritto*, 372.

'in the limelight', through the interplay of complementarity, conditioning or conflict, and therefore the judge finds himself having to compose and identify the multiple limits that pose as insurmountable boundaries to his power of decision, in an attempt to find a proper balance.

And it is precisely in the statement of reasons that one recognises the restoration of that power so broadly granted to him; indeed, to the extent to which the judge must give account in the grounds of the sentence of the criteria, indices and value judgments, the penal discretion is in that case controlled: this is because it is precisely through the statement of reasons that that control is exercised which is inseparable from the power to commensurate the penalty⁶⁷⁰. The judge must therefore demonstrate that he has reasoned and how and through what evaluations he has arrived at his final decision; it is precisely through the statement of reasons that the balance that placed two subjects on a different plane is restored and it is precisely through the statement of reasons that this balance is re-established and the parties are granted the possibility of a critical reading of his actions. Indeed, the check is carried out, as already mentioned, through the obligation to state reasons that is enshrined in general in the Code of Criminal Procedure, and also in Article 132 of the Criminal Code.

The sense of the latter autonomous provision is to reiterate and link up with the provision under Article 111 of the Constitution and, at the same time, to indicate the need for the statement of reasons to show not only the purpose, but also the progressive development of the investigation, from the fact to the personality of the agent⁶⁷¹.

There is no doubt that the obligation to state reasons is inseparable from the judge's power-duty, provided for in Article 133 of the Criminal Code, to do everything possible to try to adjust the penalty to the fact and the personality of the agent, in accordance with the purpose attributed to the penalty by the Constitution. The central point that we will try to explore further here is whether, with the criteria available, they can be considered sufficient for an assessment that covers all aspects of the accused person, and at the same time whether artificial intelligence tools accompanying the judge's assessment can be of assistance for a more complete analysis of the elements available.

Motivation represents itself as a true 'logical guardian' of certainty since it serves to determine the degree to which the judge's decision is calculable⁶⁷², in the sense that, given certain premises, if they are correct and the judge adheres to the parameters-boundaries that

⁶⁷⁰ P. NUVOLONE, *Alle soglie di una riforma*, 1563.

⁶⁷¹ F. BRICOLA, *La discrezionalità nel diritto penale*, 105 ss.; E. DOLCINI, *La commisurazione della pena*, 250.

⁶⁷² See, A. CASELLA, *Le conseguenze sanzionatorie del reato*, Turin, 2011.

govern his discretionary power, his argumentation can only lead to a given decision and its justice in terms of equality/proportionality.

Indeed, it should be possible to verify, by means of the statement of reasons, that the penalty actually imposed⁶⁷³ was arrived at by means of a logical procedure and a teleological framework that is, however, respectful of the principles, constraints and limits that guide, first and foremost at constitutional level, the judicial activity of quantifying the penalty⁶⁷⁴.

In conclusion, the duty to provide a coherent statement of reasons allows for control over the manner in which this power is exercised, in the very sense that control over the correct application of the law can undoubtedly be exercised exclusively on the reasoning supporting the decision⁶⁷⁵. It has also been observed that circumvention of the parameters of Article 133 of the Criminal Code occurs when the judge, assessing the liability of multiple defendants and for even different offences, for some of them and differently circumstantiated, for all of them indiscriminately limits himself to the generic reference 'to the circumstance referred to in Article 133 of the Criminal Code' and to the 'particular gravity of the offences and the personality of the defendants'⁶⁷⁶.

8.1 A new constrained discretion: the judge's free conviction in the face of new probative evidence. The Weakness of Articles 132 and 133 of the Criminal Code. The Praxeological Guidelines on Discretion

On closer inspection, it is noted that recourse to predictive software capable of calculating the foreseeability of the event, as outlined so far, would be technically and legally possible; however, the question of its 'effects' would remain unresolved, having regard to the principle of the judge's free conviction. Indeed, the question arises as to whether the judge could depart from the evidentiary evidence of the electronic computer or instead should consider himself bound by it. Since nothing is provided for at the regulatory level⁶⁷⁷, reference is made to the

⁶⁷³ To the assertion that the statement of reasons should be drafted in such a way as to bring out the logical path followed by the judge in his commensuration, identifying the factual criteria that guided his choices [Cass., Sez. V, 15.1.1980, Ferrero, in *Riv. pen.*, 1980, 583], followed the argument that 'the obligation to state reasons must be quantitatively correlated to the operative part' and 'the whole measure must find its justification in the statement of reasons', with the dual aim of ensuring the proper exercise of the judge's discretion and guaranteeing the defendant the appropriateness of the sentence imposed on him [Cass., Sez. I, 28.4.1987, Cardile, in *Riv. pen.*, 1988, 631].

⁶⁷⁴ Peculiar as in jurisprudence there are stances that attribute to the statement of reasons the role of 'control of reason over intuition' that doctrine has instead assigned to the provision under Article 132 of the criminal code, thus allowing the parties a critical reading of its operation.

⁶⁷⁵ In this direction, Cass., Sez. I, 27.11.1989, Andreini, in *Cass. pen.*, 1991, 1059.

⁶⁷⁶ Così, Cass., Sez. I, 14.10.1988, Balestri.

⁶⁷⁷ The 2018 Ethical Bill of Rights for the use of A.I. does not contain any principles on this point. Please refer to Chapter V.

provision of Art. 22 GDPR⁶⁷⁸ which, by prohibiting the adoption of purely automated judicial decisions unless authorised by consent, contract or the law of the Member States, would lead to the exclusion of the binding nature of the evidentiary result offered by the software. Taking its cue from and relying on such a case, the judge would then be free to weigh the various known variables and at the same time take into account the particularities of the concrete case that possibly require departing from the result of the algorithm⁶⁷⁹.

If one moves from a purely deontological level to the level of practice, one realises how the many doctrinal reminders on the inadmissibility of jurisprudential formulas, the nullity of the judgement as a consequence of the absence of motivation, and the need for conditions of transparency in the motivation, would remain 'a dead letter'.

At the doctrinal level, the actual inadequacy of the procedural instrumentation to support the judicial investigation is debated, especially for the assessment and analysis of the capacity to commit offences, which has, as has been widely criticised, a considerable lack of verifiability⁶⁸⁰; there is a risk, and it is only possible to see the effects through practice, of a so-called judicial intuitionism in making such assessments.

What comes to the fore and represents precisely the critical point on which the reasoning here is grafted is precisely the fact that if the judicial activity is entrusted with the evaluation of such elements, which then refer and innervate themselves on emotional impulses and 'intuitive lightning', the activity of determining the amount of the sanction entrusted to the judge, finds it difficult to explain and motivate itself through the instrument of motivation. The activity of determining the *quantum* of the sanction entrusted to the judge, struggles to be explained and motivated through the instrument of motivation⁶⁸¹.

Precisely on this point, a long-standing opinion of doctrine has held that precisely in the impossibility of the demonstration of the correspondence of the quantum of the penalty to the quantum of the offence to be implemented with a 'mathematical proportion' there is a risk of reducing everything through 'three or four generic standard phrases' that would serve to resolve all cases⁶⁸².

⁶⁷⁸ As will be further discussed in Chapter V. However, Article 22 of the GDPR provides that 'the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or significantly affects him in a similar way'. Thus, on this point see G. TUZET, *L'algoritmo come pastore del giudice?*, 10.

⁶⁷⁹ Per un maggiore approfondimento sul tema si rimanda al Chapter 5.

⁶⁸⁰ F. BRICOLA, *La discrezionalità nel diritto penale*, 116.

⁶⁸¹ F. BRICOLA, *La discrezionalità nel diritto penale*, 116.

⁶⁸² Così, sul punto D. MARONGIU, *L'attività amministrativa automatizzata*, Rimini, 2005, 20. The prophecy has come true: through the formulas orbiting around the appropriateness, adequacy and fairness of the punishment, in the application practice there is an obvious evasion and neutralisation of the positive discipline, on the implicit

The practice of concrete cases today reflects precisely the greater difficulties of judging bodies to reconcile the ineliminable emotional factors with the predetermination of legal constraints capable of guiding and channelling the exercise of discretionary power.

Indeed, the orientation that dominates in doctrine tends to render the judicial choice of punishment unquestionable, emphasising the intuitive-irrational components of the commensuration, so as to flatten the meaning and relevance of the obligation to state reasons on a procedural level, emptying it of substantial content and its value as a guarantee. On closer inspection, and the ground of proof is then provided by the affirmation of the admissibility of the appeal in cassation concerning the motivation of the logical procedure that leads the judge to the determination of the penalty in concrete terms is counterbalanced by the tendency to consider that the commensuration of the penalty, in the review of legitimacy, is censurable only for procedural defect (pursuant to Article 606 of the Code of Criminal Procedure) and not for erroneous application of the commensurate criteria provided by the Code⁶⁸³.

It therefore follows that, this being the case, the directives and criteria provided for in Article 133 of the criminal code lose their function of guiding constraint and run the risk of assuming only a possible and subsidiary role: indeed, in such a case the only constraint would operate only in the field of motivation and with the consequent flattening also of the provision of Article

assumption of the inability of Articles 132 and 133 of the Criminal Code to impose any substantial constraint on the judge's choices. E. DOLCINI, *La commisurazione della pena*, 66.

⁶⁸³In fact, the Court of Cassation has held that the judicial assessment of the quantum of the punishment to be imposed, being the result of a global assessment of the facts ascertained and of the personality of the offender, constitutes "an appreciation of fact that cannot be censured in the court of legitimacy", if it is congruously and logically motivated [Cass, Sec. V, 18.2.1987, Saffo, in *Riv. pen.*, 1987, 856] or an unquestionable power [Cass., Sec. VI, 6.3.1980, Coppola, in CED, 1980/145815]. Emblematic, in this sense, is the statement - now traditional - according to which, on the assumption that the judge's discretionary power must not suffer from too strict constraints, the judgement of commensuration of the penalty in concrete terms would constitute "more the result of an intuition than of a logical process of an analytical nature" [Cass, Sec. II, 8.7.1992, Pavlovic, in *Riv. pen.*, 1993, 294; even earlier, see Cass., Sec. V, 16.2.1968, Collini, in *Giust. pen.*, 1969, III, 70, Cass., Sec. II, 23.1.1980, Saponaro, in *Riv. pen.*, 1980, 886]. *Idem* for the judgement of the reduction of the size of the penalty inflicted, due to circumstances [in the case law on the merits, see C. App. Lecce-Taranto, 9.12.1994, Maiorino, in *Riv. pen.*, 1995, 925]. The emptying of the precept of the obligation to adequately motivate the choices of commensuration is induced by the emphasis on the intuitive-emotional components of the judgement; the occasional affirmation of the need for the judge to demonstrate that he has made good use of his discretionary power, so that it does not degenerate into arbitrariness [Cass, Sec. II, 2.2.1978, Di Palma, in *Cass. pen.*, 1979, 1140], does not in reality appear to be worth more than a mere statement of principle, of an ornamental nature, punctually contradicted in practice, with the impossibility of verifying a posteriori the logical path that the judge followed to concretely quantify the penalty in this or that way. Ultimately, the S.C. fosters among the judges of merit a reductive interpretation of the duty to state reasons, drastically circumscribing its sphere of operability and deliberately refraining from enhancing its guarantee purposes. For a jurisprudential stance critical of the dominant orientation in practice, see Court of Cassation, Sec. I, 28.4.1987, Cardile, in *Riv. pen.*, 1988, 631, which condemns the reference to intuitive flights of fancy, too akin to psychic impulses subtracted by their nature from any normative regulation and of an unquestionably irrational nature, having instead to be configured as a purely essential intellectual operation, governed by the canons of logic and rationality, the only ones practicable in the application of the normative prescriptions [in a similar sense, see also Court of Cassation, Sec. I, 14.9.1990, Italiano, in *Riv. pen.*, 1991, 666].

132 of the criminal code on a strictly procedural level, the real control remaining anchored only to the mere formality and regularity of the motivation. Moreover, the impossibility of asserting an erroneous application of the substantive criteria of commensuration ends up removing Article 133 of the criminal code from the scrutiny of jurisprudence, thus creating a sort of incompatibility between discretionary power and application of law⁶⁸⁴.

Thus it is as if a sort of 'free zone' were left and entrusted to the judge, which brings within his sphere of sovereignty that is entirely intangible, emptying the 'guiding' provisions of Articles 132 and 133 of the criminal code of meaning.

In this sense, the commensuration of the penalty from being a form of application of legal norms is transformed into a sphere that becomes and is left to the sovereign appreciation of the judge, inspired by equitable criteria, but with the greater risk of incomplete but also artificial motivations⁶⁸⁵.

Indeed, the tendency to import the commensurative choice on irrational and intuitive grounds, to evade the obligation to give an account of the logical process that led the judge to draw the rule in the concrete case, subtracting it from any scrutiny or verifiability, means that discretion in commensurative matters turns into free discretion or even arbitrariness⁶⁸⁶. In fact, the major problem revolves around the alleged impossibility of an ex post clarification of the logical scansion of the process of quantification of the penalty, instrumental to the recurrent affirmation of the substantial unquestionability of the commensurative choices of the judges of merit, makes it so that the obligation to state reasons pursuant to Article 132 of the Penal Code is reduced by practice to a narrow dimension with recourse to clauses of style or *tralatizial* formulas⁶⁸⁷.

⁶⁸⁴ E. DOLCINI, *La commisurazione della pena*, 58.

⁶⁸⁵ The idea underlying the pronouncements of the Court of Cassation according to which the judge's choices at the time of commensuration ontologically evade a serious duty to state reasons is resolved in the considered impossibility of expressing in rational terms the judge's evaluations and delegitimising the claim to verify whether the choice of penalty was made in conformity with the legal parameters, as if the quantification of the penalty represented the result of an unquestionable choice of the judge of merit, E. DOLCINI, *La commisurazione della pena*, 58.

⁶⁸⁶ The alleged impossibility of an ex post facto clarification of the logic of the process of quantification of the penalty, instrumental to the recurrent affirmation of the substantial unquestionability of the commensurate choices of the judges on the merits, has meant that the obligation to state reasons enshrined in Article 132 of the Penal Code has been reduced by practice to an extremely narrow and stereotypical dimension, with frequent recourse to clauses of style or *tralatizial* formulas. T. PADOVANI, *La disintegrazione attuale del sistema sanzionatorio e le prospettive di riforma: il problema della comminatoria edittale*, 320. Emblematic is the tendency of the S.C. to specify what is not necessary rather than positively clarify the content of the duty to state reasons.

⁶⁸⁷ T. PADOVANI, *La disintegrazione attuale del sistema sanzionatorio e le prospettive di riforma: il problema della comminatoria edittale*, 320.

8.1.1 *The problem of the weakness 'in the dark' of prognostic judgements*

On closer inspection, prognostic judgements undoubtedly show the elements of their weakness from the outset. Especially in areas in which prognostications are then decisive in measuring and calibrating the preventive orientation of punishment accordingly. Indeed, the object of the prognosis is largely indeterminate when it then concerns the commission of any offence.

On the contrary, when the object of the prognosis is circumscribed, more or less consciously, by the legislature, the task that is then entrusted to the judge is certainly not easier. Indeed, it is still a broad category to such an extent that an effective representation of the object of the prognosis is unlikely. In addition, there is the lack of time limits that makes the prognostic judgment so uncertain, since the very probability of reoffending in a future that has no boundaries then eludes any assessment based on the current characteristics of the defendant/convicted person; mostly if they are characteristics that are destined to change⁶⁸⁸.

Moreover, a further element that is added and that makes this type of assessment even more vague in the prognostic assessment is precisely the lack of any reference to the standard of ascertainment.

On closer inspection, the prognostic judgement is itself inherent in certain specific features. First, the theoretical structure of prognostic judgement is burdened by a 'widespread negative bias'⁶⁸⁹.

Clearly, as has already been pointed out in the preceding paragraphs, the prognostic assessment of the offender's future conduct is a judgement that poses problems and gives rise to initial reflections, firstly, from the epistemological aspect; indeed, the empirical controllability and certainty that characterise the ascertainment of the facts that have occurred are not characteristics of the prognostic judgement at the time it is formulated.

In fact, the criteria of the prognosis of social dangerousness are affected by the difficulties, uncertainties and limitations of the human sciences⁶⁹⁰.

This is an epistemological deficit that inevitably ends up conditioning, from the outset, the reliability of prognosis, both when it is concerned with the future conduct of the offender and

⁶⁸⁸ For example, think of the frequent changes in a person's social or family life context over the course of a lifetime. On the other hand, a judgment that extends over a distant time horizon can only reduce the chances of a favourable prognosis: the broader the time period, the greater the chance that the person will re-offend. Conversely, the narrower the temporal horizon in relation to which the prognosis must be made, the greater the chances that that prognosis will not be disproved.

⁶⁸⁹ In this sense, F. CAPRIOLI, *Pericolosità sociale e processo penale*, in M. Pavarini - L. Stortoni (eds), *Pericolosità e giustizia penale*, Bononia, 2013, 26.

⁶⁹⁰ See D. PULITANÒ, *Diritto penale*, Turin, 2015, 521.

when it is a matter of predicting the special-preventive effects of the punitive response. Indeed, this proven 'unreliability' does not necessarily fall entirely within the structure of the prognosis.

The first feature that undoubtedly stands out is that the ascertainment of fact, like the prognostic judgement, actually has a kind of probabilistic structure⁶⁹¹.

In fact, clearly, the higher standards of 'empirical controllability' and certainty that can be expected for the ascertainment of fact and of the criminal liability of the accused seem to depend on the greater reliability that the natural sciences can guarantee than the human sciences.

However, it is not always possible to ascertain the fact with the contribution of the natural sciences; indeed, in a large number of cases that the judge is then faced with, the natural sciences offer no help⁶⁹².

In other words, therefore, the widespread negative prejudice on this type of assessment, which weighs on prognostic judgements, should weigh equally on the ascertainment of fact.

However, what then frequently happens is that there is a sort of attitude of mistrust towards the contribution of other sciences as a support for the assessment of prognosis and therefore one ends up relegating and devolving this decision to the individual judge's own evaluation and thinking.

Underlying this attitude is probably a negative prejudice about the reliability of prognoses and a different political-criminal acceptance between this type of assessment and that of the ascertainment of the fact. Uncertainty in the latter case is less tolerated and therefore approached with greater caution with a view to full compliance with Article 27(2) of the Constitution.

Clearly, this different degree of criminal-political acceptability of uncertainty does not seem to be fully justified.

In fact, the application of the punitive response to the crime (whether it is entrusted to a penalty or to a security measure) ultimately depends on the outcome of the prognosis itself⁶⁹³.

Clearly, the deficit and the major limitation of the reliability of prognostic judgments also depend on the object of the provision indicated by the legislature. In fact, if the judge is then

⁶⁹¹ As has been observed, they are both 'fatally probabilistic, and many prognostic evaluations have a very solid rational basis'. In this direction, F. CAPRIOLI, *Pericolosità sociale e processo penale*, 26.

⁶⁹² One thinks, purely by way of example, of the ascertainment of psychic causality or of malice, which also belong to central junctures of the ascertainment of fact.

⁶⁹³ If the conviction of the innocent is intolerable, the execution of a sentence or security measure against a person whose risk of reoffending has been wrongly ascertained is equally unacceptable. The uncertainty of prognostic judgements contrasts with the re-educative purpose of punishment (Art. 27, para. 3, Const.): it would make no sense to re-educate an individual, in respect of whom an unfavourable prognosis has been erroneously formulated, but whose characteristics and needs should have led to the renunciation or change of response to the offence.

called upon to formulate a prognosis of recidivism concerning any type of offence, the criteria on the basis of which this is made will then be poorly determined and not very rational.

Therefore, the circle also appears to be closing with regard to this reflection, which thus brings the reasoning back to the starting point: the greater difficulty that is encountered today in the preparation of such instruments then necessarily leads to the identification of what are the relevant predictive factors for establishing whether the offender will then commit any crime in the future⁶⁹⁴.

A further aspect that characterises this type of judgement is a controversial aspect that has posed quite a few problems. Indeed, the question concerns whether they are abductive, inductive or deductive in nature.

According to one doctrinal orientation, the prognostic judgment would be deductive reasoning. This would differentiate it from the ascertainment of the fact of the crime and of the defendant's liability, which is instead an abductive type of reasoning.

What is certainly relevant is that in the prognosis the judge does not go in search of the best explanation of the case on the basis of the evidence gathered, nor does he use present facts to explain past facts. On the contrary, prognostic reasoning possesses a deductive nature since it "moves from the antecedent to the consequent, from cause to effect and not the other way around"⁶⁹⁵.

There is, however, a different orientation that instead maintains that prognostic judgements are characterised by the method of abduction since it is a matter of analysing a given factual situation, with the aim of formulating possible hypotheses regarding the consequences that may follow from it. Indeed, it is precisely through abduction that the judge can then formulate hypotheses on the future. In contrast, in prognosis, the judge finds himself using the abductive method of inference towards the best explanation; he should thus formulate hypotheses that appear reasonably possible on the basis of the elements of knowledge and indications available to the judge at the time he formulates the prediction, in order to then be able to exclude hypotheses on future events that conflict with the available evidence.

⁶⁹⁴ It is very peculiar how this assessment is made in the English system; indeed, even here it is interesting to note the different experience offered by the English system. As we have had the opportunity to elaborate, the boundaries of the object of the English prognostic judgement are well drawn through reference to two different factors. The English judge, in fact, on the one hand has to foresee that the perpetrator will commit one of the serious offences contained in a list identified by the English legislator (contained in the Criminal Justice Act), on the other hand he must also establish that serious harm to the victim may result from the commission of that offence.

⁶⁹⁵ Thus, on this point, F. CAPRIOLI, *Pericolosità sociale e processo penale*, 23; in the same direction C. ZAZA, *Il ragionevole dubbio nella logica della prova penale*, Milan, 2008, 44 ss.

However, the element on which we certainly do not take a favourable position concerns the fact that deductive reasoning does not seem to follow the logical model to which prognostic judgments are to be traced. Indeed, in deductive reasoning, if the premises are true, then the conclusion must also 'necessarily be true'.

However, and this, if only partially, is the crux of the matter to be resolved, it does not appear that in the case of prognostic judgements, the existence of true premises can logically imply the conclusion (as to the offender's future conduct and the effects of the offence's response).

In the prognostic judgment, indeed, there seems to be no trace of the soundness, certainty and wide margins of certainty of deductive inferences.

In conclusion, prognostication would thus appear to be in the nature of inductive reasoning since it moves and starts from premises about objects that have been examined to conclusions about objects that we have not examined'.

Lastly, therefore, prognostic judgements on recidivism and the effects of responses to the offence would appear to be based on a probabilistic inductive inference that has been consolidated on the basis of previous empirical experience: given certain personal characteristics, it is then assumed that that person will behave in the manner observed in previous cases that had the same characteristics (as starting from an unknown fact).

Indeed, it is noted that the more frequently a causal connection between certain personal characteristics and reoffending is observed, the more solid the inductive inference will then be, even though it is always logically possible that the behaviour will be different from that previously observed.

8.2 The concept which returns: social dangerousness within Article 133 of the criminal code.

Indeed, more appropriately, if one were to take the provision under Article 133 of the Criminal Code, one should not so much speak of social dangerousness as of the offender's capacity to commit a crime. A provision that looks at and considers within it a large container with several elements: the assessment of personal, anamnestic, behavioural and contextual aspects. Indeed, some refer to and are closely connected with the offence committed, while others are independent of it⁶⁹⁶. It follows that the assessment of dangerousness must be carried

⁶⁹⁶ Examples of criteria for evaluation include the difference between a behaviour of lucid indifference and complacency that may signal a particularly strong criminal inclination, as well as trial conduct that denotes obstinacy or insensitivity to the victim and accomplices. On the contrary, immediate repentance, hesitancy and

out through the full recognition of all factors relating not only to the seriousness of the offence but also to the offender's capacity to commit crimes. Indeed, the criteria identified by the legislature are the same as those laid down for determining the penalty. However, it is clear that the factors concerning the offender's capacity to commit offences, analysed from a prognostic point of view⁶⁹⁷, may have a different meaning from what is assumed instead from a merely retributive point of view, depending on the fact that the offence committed is not really considered as such, but as a symptom of probable future recidivism. Indeed, the circumstantial elements of dangerousness, relevant to the offender's capacity to commit offences, are, pursuant to Article 133 of the Criminal Code: the offender's motives to commit offences and his character; criminal and judicial precedents and in general the offender's conduct and life prior to the offence; conduct contemporaneous with or subsequent to the offence; the offender's individual, family and social life conditions.

On closer inspection, when the judge is called upon to determine the type and measure of the penalty to be imposed, he must take into account the future conduct of the offender. Indeed, among the criteria for the commensuration of the penalty provided for in Article 133(2) of the Criminal Code, the offender's capacity to commit offences is central. However, the legislature has not provided a clear definition of this concept but has however identified several factual criteria on the basis of which the judge must reconstruct the offender's capacity to commit offences. Within this assessment, the latter must consider several elements, such as: the personality, the conduct of the subject (prior, contemporaneous, subsequent to the offence), the living conditions (individual, family, social) of the offender. While it is true that there is no agreed view of what is to be understood by capacity to commit offences, the interpretation that is preferred, however, would seem to be that which identifies this concept in the possibility and capacity to commit further offences in the future.

This premise, therefore, in view of the constitutional dictate dictates that if punishment must therefore meet the re-educational requirements⁶⁹⁸ laid down at constitutional level, at the same time it is necessary to take into account the specific characteristics of the individual.

collaborative trial conduct are commonly regarded as symptomatic indications of a not so firm criminal inclination.

⁶⁹⁷ Any prognostic judgement, based on the appreciation of the recurrence of a 'danger', which is by its very nature a judgement addressed to the future, which excludes its possible declination in terms of historical certainty (an attribute by which one can, conventionally and procedurally, qualify only past conduct), with an ineradicable margin of fallibility', see A. M. MAUGERI, *L'uso di algoritmi predittivi per accertare la pericolosità sociale*, 8.

⁶⁹⁸ On this point G. FIANDACA - G. DI CHIARA, *Una introduzione al sistema penale. Per una lettura costituzionalmente orientata*, Naples, 2003, 40 ss.; nonché E. GALLO, *L'evoluzione del pensiero della Corte costituzionale in tema di funzione della pena*, in *Giur. cost.*, 1994, 3204.

The problem that arises, however, is given by the fact that there are several factors that certainly do not make such an assessment easily achievable and, at the same time, make it imprecise or dictated by intuition and convictions.

In fact, starting precisely from the premise that these are rather difficult evaluations since they are dictated by various elements combined, it is true that the more they are considered as a whole, the more likely it is that they can correspond to a perhaps more precise assessment of the future.

In making this premise, the first difficulty is certainly centred on the fact that the judiciary and the problem of the efficiency of the quantity of loads and cases to be followed often do not allow for a careful analysis not only of the file, but also of the individual elements to be considered in the concrete case; at the same time, even when the individual judge has the possibility (and availability) of being able to manage and follow a given case carefully, many times, in some cases, the evaluation of the same cannot be completely complete and centred on all the elements to be taken into consideration. This is also because the judge himself has a limited appreciation and also the human mind is not always able to grasp all elements and compare them with each other. The *surplus* and the decisive element that certainly distinguishes it from a machine are given by the fact that the capacity and singularity of the human mind is able to grasp the different nuances given from case to case and thus make differences at the same time. This capacity, therefore, as we will attempt to outline at the end of the paper, immediately brings to light what could therefore be the possibility of improving the evaluation by allowing within it the support or aid of an evaluation issued by a machine alongside the evaluation derived from the human mind⁶⁹⁹.

8.3 *Problems and the first emerging evidence on the phenomenological level*

On closer inspection, not only the praxeological but also the phenomenological level present critical profiles that unfold in the evaluative operation.

In the first place, the role of legislative power assumes considerable importance: in fact, the gaps that the legislative criteria of the commensuration of punishment present are exploited to correct downwards the anachronistic legislative evaluations expressed in the edictal spaces⁷⁰⁰. Moreover, secondly, the intrinsic inexpressibility of the motivations or arguments, creates and gives rise to a real dissociation between the formation of the judgement on the measure of the

⁶⁹⁹ *Segue*, Chapter V.

⁷⁰⁰ Così, sul punto, L. MONACO, *Prospettive dell'idea dello "scopo" nella teoria della pena*, 283 s.

penalty and the relative motivation; often, such factors conceal the recourse to factors or parameters extraneous to Article 133 of the criminal code⁷⁰¹.

As a matter of fact, and this is one of the weak points of the discretion entrusted to the judge in the choice of the commensuration of the punitive treatment, recourse is made, as is well known, in practice to extralegal and hidden criteria of commensuration; indeed, the reference to tradition as a factor of stability and certainty, the comparison with previous judgments, also risk crystallising a sort of so-called 'penalty tariffs'. Other factors that undoubtedly come into play are the emotional impulses of public opinion or the sanctioning customs in the various judicial districts; all factors that, however, risk leading to unreasonable inequalities in treatment⁷⁰². Parallel to this phenomenon, there is also the 'emptying out' of the duty to state reasons and the consolidation of a true 'judicial penal policy'.

It is necessary to assess, over and above the judge's obligation at the procedural level to give reasons for his decision, how the content of the obligation to provide a specific statement of reasons regarding the criteria justifying the judge's use of discretionary power is understood in the practice of application, and what are considered to be the ways in which this obligation is fulfilled. The datum that emerges from the case law is undoubtedly that of the persistent tendency to evade and minimise the significance and importance of the obligation to state reasons, through the almost total reduction of analyticity and specificity. Indeed, the techniques of simplification of the commensuration of the penalty adopted, made pressing by the quantitative dimensions of the phenomenon, are accumulated by the need to then aggrate the rule under Article 132 of the criminal code⁷⁰³.

On closer inspection, it would seem appropriate to make a brief and precise reflection on the nature of the judgement of commensuration of the penalty because if it is true that the same is the result of an intuition deriving from a global assessment of the facts as ascertained and of the personality of the offender, it is considered at the same time sufficient to take into

⁷⁰¹ Indeed, what is most worrying, as has been sharply pointed out in doctrine, is that 'the protagonists of the commensuration of punishment in our system continue to be tradition, irrational factors and the criminal policy views of individual judges', thus on the point, E. DOLCINI, *La commisurazione della pena*, 58.

⁷⁰² *Ibidem*, 42.

⁷⁰³ Indeed, on the assumption that the judgement of commensuration is the result of an intuition deriving from a global assessment of the facts as ascertained and of the personality of the offender, the taking into consideration, albeit implicitly, of the elements indicated in Article 133 of the criminal code is considered sufficient. Faced with a range of discretion as vast as that entrusted to the trial judge by the combined provisions of articles 132 and 133 of the criminal code, the inexistence of a statement of reasons explaining the reasons for the differences between the amount of the penalty concretely chosen and another slightly lower (or possibly higher) one is affirmed: the obligation must be considered to have been fulfilled every time the trial judge's choice falls on a penalty that, in terms of its overall amount, does not appear manifestly disproportionate to the fact being punished. The concrete determination of the penalty is therefore seen as the result of an overall assessment, and not of an analytical judgment on the various elements offered by the law.

consideration, albeit implicitly, the elements and criteria indicated in Article 133 of the criminal code.

The biggest problem arises because, faced with such a wide range of discretion as is entrusted to the trial judge, who remains anchored only to the two provisions of Articles 132 and 133 of the Criminal Code, the inexistence of a statement of reasons explaining the reasons for the differences between the amount of the penalty concretely chosen and another only slightly lower is affirmed. In fact, the obligation to state reasons should, on the other hand, be considered fulfilled whenever the choice of the trial judge falls on a penalty that, in terms of its overall amount, does not appear manifestly disproportionate with respect to the fact that is the subject of the penalty⁷⁰⁴. Thus, even the concrete determination of the penalty is seen as the result of an overall assessment and not of an analytical judgment on the various elements offered by the law⁷⁰⁵.

9 *The human decision and the technological decision: a surmountable opacity? An adversarial 'technicalisation'*

On closer inspection, one of the major critical profiles that is posed and that has led to the approach or reflection on the possible coupling of an artificial intelligence tool to the judge is precisely that connected to the 'external' controllability by the judge of the result of the predictive software. Indeed, it is precisely the principle of cross-examination that represents, in the trial context, the 'natural antidote' against the tampering and instrumentalisation of evidentiary results since it allows for a real confrontation between the parties in their assumption and formation⁷⁰⁶.

Indeed, it should be noted from the outset that the admission of such instruments must necessarily be confronted with the moment of cross-examination, as a procedural and substantive guarantee. The cross-examination is guaranteed, first and foremost, at the time of the admission of evidence (as provided for in Rules 493(1) and 496(1) of the Code of Criminal Procedure); in fact, the judge, having heard the parties, admits atypical evidence or scientific evidence requested, if suitable for ensuring the ascertainment of the facts. Further decisive

⁷⁰⁴ Cass., Sez. I, 27.11.1989, Andreini, in *Cass. pen.*, 1991, 1059].

⁷⁰⁵ Cass., Sez. VI, 20.5.1989, Mancusi, in *Riv. pen.*, 1990, 335; on the non-necessity of an analytical statement of reasons and the sufficiency of an overall assessment of the elements of Article 133 of the Criminal Code, see Cass., Sez. II, 26.3.2008, Gasparri, in CED, 2008/239754; Cass., Sez. V, 21.9.1982, Urtoller, in *Riv. pen.*, 1983, 533.

⁷⁰⁶ Precisely on the subject of the relationship between scientific and adversarial evidence, see P. TONINI, *Prova scientifica e contraddittorio*, in *Dir. Pen. Proc.*, 2003, 1459, which emphasises the need for the adversarial principle to be implemented also in the formation of scientific evidence.

moments concern the phase of the taking of evidence and above all, in the course of the preliminary investigation, the examination and cross-examination of the expert, who may be called upon to answer for the result of the software, its interpretability, any margin of error, whether ascertained or potential, the reliability of the system, the design method, the data entered, etc.

The transparency of the predictive scientific method, which is the basis of the computational model, should be guaranteed, through "a more meaningful discovery mechanism, providing for direct confrontation between the experts appointed by the parties and allowing direct questions between them"⁷⁰⁷.

On closer inspection, the question of the guarantee of cross-examination becomes very delicate in cases of privately owned software since the algorithm is covered by copyright⁷⁰⁸.

This last-mentioned case acted as a watershed between the time when these instruments had silently entered into application and, on the other hand, the time when the first questions concerning the possible negative effects were beginning to become apparent. This case in fact helped to highlight the problem of the opacity of the decision-making processes of private algorithms. Indeed, in the context of criminal proceedings for the offences of driving a vehicle without the owner's consent and attempted violation of a checkpoint, the Wisconsin Supreme Court on the one hand confirmed the legitimacy of the use of these predictive computer tools and, on the other hand, addressed the issue of transparency and the reviewability of the final result by the judge⁷⁰⁹. However, the Supreme Court on that occasion rejected the appeal, ruling

⁷⁰⁷ On this point, O. DOMINIONI, *La prova penale scientifica*, 1063.

⁷⁰⁸ The question had arisen in the leading Loomis case of 2017. In the decision Wisconsin Supreme Court, *State v. Loomis*, c. 2015AP157-CR, 13/07/2016, in *Harvard Law Review*, 2017, vol. 130, 1530 ss. Così in commento anche, S. QUATTROCCOLO, *Quesiti nuovi e soluzioni antiche?*, 1748 ss. In this decision, the case concerned an individual, Eric Loomis, who had been convicted of the offences of driving a vehicle without the owner's consent and attempted violation of a checkpoint. For the purpose of determining the sentence, during the investigation, the prison police officer had produced a PSI report, containing the results of the recidivism and social dangerousness risk assessment carried out with the Compas software. In particular, this tool is based on a twofold source of data: on the one hand, on information obtained directly from the defendant, in an interview given to him through a pre-established questionnaire, and on the other hand, on the certificate of criminal records and pending charges; this information is then processed through a computational model in relation to control statistical data, referred to a sample of the population. At the final hearing, the Court, having examined the results of the risk assessment, which showed a particular proclivity to commit crimes, sentenced Loomis to a particularly high sentence (six years imprisonment and five years of special surveillance, the so-called extended supervision). However, given the obvious disproportion between the sentence imposed and the seriousness of the crime, the defence proposed to the trial judge a motion for post-conviction relief, at the same time contesting several aspects of the violation of the due process principle and highlighting some critical aspects related to the use of the risk assessment tool in the deliberative phase of the sentence.

⁷⁰⁹ In particular, the defence had first of all contested the impossibility of checking the operating mechanism of the Compas software, used in the present case for calculating recidivism and assessing the agent's social dangerousness (since it was covered by trade secret). It is also interesting to note that the instrument in question has been subjected to a number of practical experiments (not all of them independent) whose results have always seemed decidedly contradictory. Indeed, in the face of some partly positive conclusions, some studies have

out the violation of due process rules, given the possibility for the defendant to compare individual starting data (input) and final risk assessments (output) on the basis of the software's user manual. However, as a result of this comparison, the defence noted that the programme did not provide completely objective results, since it took into account, in a manner unfavourable to the defendant, some data of 'Lombrosian memory', such as male gender, ethnicity and even somatic characteristics, which should be considered 'neutral, on pain of violating the right to equality, the right to be sentenced on accurate information and the right to individualised treatment (the so-called right to an individualised sentence). d. right to an individualised sentence). In order to overcome this criticism, the Supreme Court held that the individual rights of the convicted person were not violated in this case, since the results of the software were only used 'in support' of the judge's decision. In conclusion, in the Loomis case, the Court decided to save the use of such tools at least at the sentencing stage, for the purpose of determining the sentence, despite the fact that several problems concerning the opacity and lack of transparency of the copyrighted decision-making mechanisms had been complained of.

Indeed, having said that, such a scenario would appear from the outset to be incompatible with the guarantees of cross-examination, the right of defence and due process provided for in Article 111 of the Constitution.

It follows that, where the transparency of the decision-making mechanism cannot in fact be guaranteed, recourse to algorithmic evidence should be excluded in order to avoid the risk of a "black box decision".

9.1 The algorithm in the decision-making phase: what benefits and towards what future?

Undoubtedly, evidence-based decision-making is part of a broader pattern in contemporary society that in itself implies the use of scientific research to improve the quality of decision-making. Indeed, as in several fields, science offers empirically grounded guidelines for the

highlighted risks of discrimination and low significance, in terms of the risk of recidivism, of the parameters used. In particular, a study conducted by the American NGO ProPublica took up precisely the latter profile, of the criminogenic irrelevance of certain factors used by the model, causing quite a stir in public opinion. Indeed, ProPublica performed an evaluation of the data broken down by ethnic groups and ascertained that the algorithm is particularly fallacious in labelling young blacks as future criminals, possessing an error rate twice as high as that of Caucasians. Furthermore, the normalised analysis showed that black defendants were approximately 77 per cent more likely to be labelled at increased risk of committing future violent crimes. Thus, on the point, S. QUATTROCOLO, *Quesiti nuovi e soluzioni antiche?*, 1748 ss.; L. MALDONATO, *Algoritmi predittivi e discrezionalità del giudice: una nuova sfida per la giustizia penale*, 401 ss.

exercise of criminal justice, as part of a gradual trend towards the use of evidence-based practices in law⁷¹⁰.

This should also identify the individualised treatment that best meets the offender's needs. In particular, the idea of introducing an 'individualised, evidence-based programme tailored to each defendant's risk factors and needs', in order to put in place policies even in the post-prison phase, truly aimed at preventing re-offending in the knowledge that 'evidence-based sentencing merges punishment with rehabilitation. Imposing a sentence with appropriate conditions based on the defendant's individual risk to reoffend and need for treatment or programming does not diminish the prosecutor's role in advocating for appropriate sanctions: it makes the criminal justice system more effective'⁷¹¹.

Therefore, the proposal to introduce systems based on evidence-based practices in the Italian legal system presupposes, as a first element, the necessary identification of the so-called risk 'factors' or 'risk predictors'⁷¹² that are directly involved in the criminal behaviour and that are very relevant and necessary to carry out the analysis and assessment of the dangerousness of an individual⁷¹³.

As already mentioned, this type of factors can undoubtedly allow an actuarial (or statistical) approach to the assessment of criminal dangerousness.

10 The paradigmatic value of Article 133 of the Criminal Code: inadequate criteria?

Indeed, a fundamental value is recognised in Article 133 of the Criminal Code, which represents, in the Italian legal system, the central part on which many other provisions depend⁷¹⁴. This case performs precisely the function of a hinge between traditional criminal law, based on the 'imputability-penalty' binomial, and what is instead called *novum organon*, which is inspired instead by the 'social dangerousness-security measure' binomial.

⁷¹⁰ See, S. D. HART, *Evidence-Based Assessment of Risk for Sexual Violence*, 145-46; R. E. REDDING, *Evidence-Based Sentencing: The Science of Sentencing Policy and Practice*, in *Legal Studies Research Paper Series*, Paper No. 09-41, 2; L. D'AGOSTINO, *Gli algoritmi predittivi per la commisurazione della pena. A proposito dell'esperienza statunitense nel c.d. evidence-based sentencing*, 359.

⁷¹¹ K. HEILBRUN (et oths), *Risk-Assessment in Evidence-Based Sentencing: Context and Promising Uses*, 1 *Chapman J. Crim. Just.*, 127, 2009, 10135 ss.

⁷¹² On this point, see Chapter 5.

⁷¹³ Factors that may, inter alia, relate to: age, gender, ethnic origin, level of schooling, family and work situation, social position, previous criminal record, previous prison experience, places and people frequented, presence of offenders in the family circle or network of acquaintances, place of residence, difficulties in regulating anger and aggression, impulse control, a history of previous acted violence, a history of hospitalisation, pro-criminal thinking, certain contextual variables (such as, for example, lack of family and social support), drug or alcohol consumption, psychopathy. Così F. BASILE, *Intelligenza artificiale e diritto penale*, 17.

⁷¹⁴ Così sul punto, G. GUARNERI, *Pericolosità sociale*, 867.

From the analysis and development of the paper, the main features inherent in certain types of prognostic judgements in the Italian legal system certainly emerge. Indeed, for the prognostic judgement and, more generally, for the entire phase of commensuration of the penalty, it is necessary to enter the most vulnerable and private sphere of the offender's person.

However, as is well known, within a criminal law of the fact, there is a criminal procedure aimed at verifying an accusatory hypothesis and, consequently, an evaluation by a judging body that must also be able to make assessments that look to the future.

At the theoretical level, clearly, in order not to run the risk of a 'criminal law of authorship', the author of the fact always remains in the background: only the personal characteristics that are functional to the ascertainment of the fact and liability are then subject to assessment. Much of the information relevant to the formulation of the prognostic judgment and, in particular, that mentioned in Article 133(2) of the criminal code.

In particular, precisely in relation to the capacity to commit offences rarely constitute the subject of ascertainment in a trial of cognition⁷¹⁵.

In conclusion, precisely in the face of the crisis and uncertainty of certain types of assessments, the opportunity arises to avoid or at least diminish improvisation, subjectivism, the exclusive domain of intuition and to rely instead on prognosis formulation models that are methodologically rigorous and verifiable. This is an objective that has perhaps often been neglected more for the commensuration of punishment than for the study of the application of security measures. For this reason, here, after having primarily presented and described the current application methods of certain instruments, an attempt is made to propose, even if only in a purely theoretical key, the possibilities and findings of application⁷¹⁶.

⁷¹⁵ In proceedings in which the defendant proclaims his innocence and has the right not to make self-incriminating statements, it would make no sense for him to reveal information, albeit relevant to the prognostic judgement, that could be used to his detriment in the assessment of responsibility for the act. The defendant has no interest in making statements on his individual, family and social life conditions or on his conduct prior to, contemporaneous with or subsequent to the offence (Article 133(2) of the criminal code), which might also be relevant for the formulation of a favourable prognosis, if it can be inferred from such statements that the defendant committed the offence.

⁷¹⁶ The Anglo-Saxon approach is different: there, about 6 out of 10 experts make use of one of the 400 or so risk assessment tools available, in order to enable the judge to make a prognostic judgement on recidivism based on scientifically reliable cognitive data; thus, G. ZARA – D. FARRINGTON, *Criminal recidivism: explanation prediction and prevention*, 148 ss.; in relation to "the need to subject inferences to controls as suitable as possible to ensure at least a sufficient degree of reasonableness and reliability" to prognostic judgements M. TARUFFO, *Sui confini. Scritti sulla giustizia civile*, 340 s.

10.1 *The ethicality of human judgement and its ineradicable subjective components*

Indeed, as emerged from this initial analysis conducted in the first chapters, we take note of how human beings instinctively place high expectations on the tools of technology and advanced artificial intelligence. As already mentioned, the myth of digital justice based on pure mathematical calculation has always had a strong and undoubted suggestiveness, pursuing the utopian as well as dystopian idea of a more 'exact' justice.

Indeed, by means of objective and impartial calculation systems, such because without the human component, one is fascinated by the idea that it is a possible future that supplants and overpowers the fallibility of man, conditioned by prejudices and other imponderable subjective variables.

In recent years, on the opposite side, neuroscience itself has helped to highlight the limits of human rationality and critical thinking⁷¹⁷.

On closer inspection, the 'technocentric vision' would risk losing the importance and centrality of the 'more human' component, that competent value, ethical and spiritual component of the judicial decision, of the so-called human face of justice. However much one thinks one can bring the machine as close as possible to the idea of a human being, however, there remains that insurmountable limit of free, critical and authentically creative thinking that would remain inimitable and irreplaceable.

The characteristics that make the *humanitas* of thought irreplaceable derive from those very characteristics that make it sometimes fallible and sometimes (perhaps) 'not right'.

In imagining then perhaps the overcoming of human and digital justice, one might think of keeping them two separate entities, with the human one using the digital one to 'improve' outcomes.

This paper precisely pursues the objective of overcoming indiscriminate closures to the subject and, at the same time, maintaining an open stance to possible forms of interaction, coexistence and collaboration between human decision-makers and digital programmes.

In this sense, one could imagine how the development of integrated digital justice could prove functional in overcoming the criticalities and limitations of human evaluation⁷¹⁸.

In conclusion, the time is probably not yet ripe in the Italian legal system to give entry to predictive algorithms in the assessment of the criminal dangerousness of a defendant in the

⁷¹⁷ On this issue, F. BRACCO, *L'Homo errans nell'era dell'infalibilità tecnica*, in *Altre Modernità*, 263 ss.

⁷¹⁸ D. PERRONE, *La prognosi postuma tra distorsioni cognitive e software predittivi. Limiti e possibilità del ricorso alla "giustizia digitale integrata" in sede di accertamento della colpa*, Turin, 2022, 120.

function of supporting the decision of the human judge, but the US experience solicits reflections on the future scenarios that may loom on the horizon of the legislature before which the latter cannot be found unprepared. Only with a use of algorithmic systems of riskassessment that is normatively disciplined and above all 'constitutionally oriented' will it be possible to affirm that the rights of liberty and defence celebrated by the Constitution have not yielded before the inexorable progress of information technology.

11 Concluding remarks: drawing conclusions on risk assessment

On closer inspection, the analysis carried out in the first two chapters was necessary in order to provide information on the leading instruments involved in this investigation.

Certainly, in the first place, one becomes fully aware of how numerous and varied are the prognostic evaluations that are referred to the judge in his daily life: in other words, one wonders to what extent this concept of dangerousness actually occurs, which does not only affect the cases expressly required by certain institutions (one thinks, for example, of precautionary measures, prevention measures), but even the commensuration of the sentence.

Having passed this first stage, one realises that tools such as risk assessment with certain characteristics and peculiarities of artificial intelligence systems come to the fore in this type of processing.

Then, reflecting on the application boundaries of such tools and at the same time of the possible legal consequences, one realises how such risk assessment tools, as models that in themselves convey psycho-criminological theories, cannot escape the evidentiary rules applied by the courts for the admission of scientific evidence. Each legal system, in fact, will have the freedom to decide, within its own discipline on evidence, what the boundaries are for admitting or excluding it at the trial, decisional and precautionary stages, in the event that it does not present the minimum requirements that are required by law.

Lastly, it was decided in the last chapter to consider the aspects inherent in the regulatory framework and how it intervenes in this proposal and, at the same time, to refer to the proposition of the relevant issues pertaining to the applicability of these instruments (related to all the most discussed profiles).

In conclusion, it is noted that, given the gaps presented and described in our legal system and the lack of objective indices and criteria present in this type of assessment, if the prognostic instruments are not refined, the individualisation of punishment would remain little more than

a mere political-criminal wish and the catalogue of penalties would become merely a tool whose effects and special-preventive efficacy are unknown.

Chapter Five

Perspectives de iure condendo: the feasibility of a proposal.

Comparing rights: the difficult balancing act between new rights and previous guarantees.

A concluding critical analysis

SUMMARY: 1. Methodological premise: towards the proposition of a model and a look at the past. – 2. Future perspective: the proposed model. – 2.1. The output of the algorithm: predictions. – 2.1.1. At which stage could predictions intervene? – 3. An initial critique of accuracy and the risk of generalisations. – 4. Risk indicators: the difficulty of selection. – 4.1. The most sensitive issue: a chance to overcome. The choice of risk factors. – 5. The forward-looking perspective and its structure: individualised judgement and human control of the judge – 5.1. A key to begin with: the algorithm applied only *in bonam partem*. – 6. Comparing rights: the feasibility of a proposal between a balance of rights and guarantees. – 6.1. The narrower frame of applicability: the possibility of intervention in the face of relevant factors with dynamic characteristics. – 7. The main issues arising from the first reflections. – 7.1. Constitutional limits. – 7.2. Compression of personal liberty: between rights and balances under Article 5 ECHR. – 7.3. The risk of profiling: the use of big data and the invasive approach with individuals. Article 22 GDPR and its regulatory boundaries. – 7.3.1. The collateral risks associated with profiling: the delicate drifts of stereotyping. – 8. The problem of discrimination overcome by the rationality of the machine? – 8.1. The problem of data transparency and the opacity of A.I. systems. – 8.1.1. Current *scenario* and possible future solutions. – 8.1.2. The risk of undermining the guarantee of the 'equality of arms at trial': the right of access for the defendant. – 8.1.3. A possible solution to overcome the obstacle of obscurity: towards greater transparency. – 9. The necessity and clash of penal guarantees. – 9.1. The difficult balance between presumption of harmlessness and presumption of innocence. – 9.1.1. The risk of determinism in decisions. – 10. The regulatory framework: the regulatory sources of Artificial Intelligence. – 11. New perspectives and positions on artificial intelligence: the EU White Paper. – 12. The Ethic Charter of EU. – 13. The Council of Europe's position on automated decisions with profiling. – 14. Regulatory limits to artificial intelligence: ethical and legal barriers? – 14.1. The General Data Protection Regulation. – 14.2. Automated data processing: a step forward to the GDPR with Legislative Decree No. 51 of 2018. – 14.3. EU Directive 680/2016 on the processing of personal data for the prevention, investigation, detection and prosecution of criminal offences or the execution of criminal penalties. – 15. The proposal for a future regulation on A.I. *de iure condendo* perspectives. Progress towards regulation?

1 Methodological premise: towards the proposition of a model and a look at the past

The preceding chapters have attempted to provide a general overview of the current state of the art and, at the same time, a description of the technologies that are considered as protagonist of criminal justice today.

Upon close examination, in this dialogue between technologies and 'ancient' systems, one must consider the emergence of a fateful 'clash' in this most topical challenge of *technicalising* Criminal law.

On closer inspection, the descriptive and argumentative analysis of the first four chapters undoubtedly highlights a new panorama: the emergence of new technologies confronting criminal law today. It is a 'confrontation' that has seen in other legal systems the inclusion of new instruments that have in part made it possible to 'rethink' certain institutions or simply

types of 'decision' or human decision-making, in light of new protagonists. Clearly, before arriving at the possible proposal of some of these in the (still) imaginary insertion in the criminal trial, it was first necessary to ask ourselves what they are, how they work, and which of these instruments we have already seen used, albeit in other jurisdictions. In a second moment, we have tried to identify the most delicate phases in which a judge has to make prognostic evaluations, in which it is believed that these instruments can assume a leading, or at least supporting, role.

This last chapter will therefore draw conclusions and propose one of the models described in the preceding pages. In doing so, however, the analysis must take into consideration, albeit at a later stage, those rights and guarantees with which the introduction of such instruments 'clashes'.

In this concluding part, therefore, an attempt will be made to frame in greater detail the future perspectives to be presented in the Italian process and the possible introductions within it. An attempt will also be made, in the central part of the chapter, to bring together the major critical issues and questions that arise on the subject and, at the same time, the possible solutions that can be envisaged.

Therefore, an attempt will be made, in a more defined and in-depth manner, to "pull the sums" on the proposed scenarios in Italy. In doing so, this paper will illustrate the solution or proposal examined in this elaboration and compare the sources (including soft law acts) that protect artificial intelligence and, at the same time, the fundamental principles and guarantees at a penal and constitutional level for which the A.I. must be either reconsidered or readjusted.

Indeed, as already anticipated, during criminal proceedings, the judge is called upon, according to different declinations, to make different predictive judgements on the dangerousness of the defendant. Therefore, the proposal, albeit at a theoretical level, concerns the possible introduction of A.I. instruments in the most delicate phase of the judging body, while attempting to balance the rights and guarantees at stake and the possible advantages of the proposed application of these instruments.

Lastly, while taking a very cautious stance, in the course of the study, a very important aspect was revealed: the fallibility of the judge is a theme that has always interested the minds of jurists, but, even more specifically, in this case, the idea that there are certain evaluations that cannot only make use of the fact that has occurred, but that turn to the past to make a decision that concerns the present and the future. These are precisely the prognostic evaluations that by their nature and characteristic represent a greater degree of fallibility and uncertainty.

As we have seen in the analysis in Chapters 2 and 3, the legislature is not always able to provide the adjudicating body with limits and, above all, criteria to which it must adhere when making its assessment. It follows that in many cases such assessments have an even more pronounced component of subjectivity on the part of the judge as he is looking to the future. Moreover, the sector studies would seem to show how the actuarial assessment of the offender's risk of reoffending can be much more precise than the human one, since it can process an immense amount of data of which no judge could reasonably dispose. In particular, the use of the so-called mixed-type assessment tools - designed to give relevance not only to the statistical findings but also to the examination of the offender's personality - could offer a useful guide to direct the judge's activity in the commensuration of punishment. To this end, the concluding part of the contribution will assess the possibility of replacing the current model of pure judicial discretion with that of conditional discretion, in which a significant role could be attributed to predictive algorithms and, in general, to methods of actuarial assessment of the offender's capacity to commit offences.

Having said this, although we started in the analysis of this study from a much more cautious and 'fearful' perspective, it is now considered that precisely one of the most important and even complex human activities, which has as its object the most precious asset of personal liberty, should not be evaluated. At the same time, it cannot be thought the same, and in its more complex declinations such as predictive judgment (necessary and ineradicable), can be excluded a priori - by virtue of an absolute and a priori presumption of unreliability - with the help of science and technology. At the same time, the intention is not to propose any substitution of human activity, except for support of the judicial decision only at certain specific moments, and then to try to balance its criticalities and identify corrective measures to adapt these new subjects to an order that stands with its guarantor system and the rule of law.

In conclusion, it seems necessary to reiterate that the need for a very cautious reflection on the current issue is due to a problem that is undoubtedly to be found in the Italian legal system. As already mentioned, the greatest obstacle is to be found in the various difficulties encountered by the judging body when it finds itself having to make a prognostic judgement. This is because the legislature has formulated it in significantly vague terms (e.g. deciding whether the offender will refrain from committing other offences in the future) without clearly

and more precisely circumscribing the boundaries of the prognostic judgement, either in terms of time or according to the type of offence⁷¹⁹.

As noted during the discussion of this paper, there is a further element of complexity concerning this difficult relationship between knowledge and scientific notions, and concepts proper to legislation and, in this case, proper to criminal law. In fact, concepts such as 'social dangerousness' and 're-education of the offender', which would otherwise remain completely inaccessible to the expert called upon to support the judge in predicting the offender's future behaviour, are translated into an almost necessary 'scientific knowledge'.

A final problem which, however, will not be dealt with here as deserving separate treatment, relates to the element of the margin of error concerning prognostic judgments. Indeed, in this case, the question arises as to whether the assessment must be made on the basis of the expected standard of 'beyond reasonable doubt' or, on the other hand, whether the 'preponderance of the evidence' and, consequently, the permanence of possibly reasonable or probable doubts on the future conduct of the offender is sufficient.

To attain a better awareness of the phenomenon and the proposed perspectives, an attempt will be made in this last part of the paper to conduct a reasoning of the risk and regulatory limits of the entry of predictive algorithms within the Italian criminal trial. This last part of the analysis will be conducted by attempting to assess and identify the regulatory barriers (but not only) that would prevent their entry.

2 *Future perspective: the proposed model*

On closer inspection, the model that we wish to propose here is a supervised machine learning model. Specifically, it is an algorithm (learning) or also called a training algorithm (training) that is able to learn and consequently build a model, using the tests provided to the system and the algorithm or model learned.

In particular, as described in the first part of the paper, the algorithm uses the training set to construct a model of the task to be performed by the system. Such a model may be seen as a mathematical function, i.e. a mechanism that links possible inputs to corresponding outputs. For instance, the model could link possible images of animals to words denoting the

⁷¹⁹ Clearly, problems do not arise and arise only because of a problem of legislative rules. But further factors further complicate the formulation of the prognosis on the risk of reoffending because they add unpredictable environmental and situational factors of external conditioning and the time period (which may consequently be more or less extensive) of validity of the prediction on the offender's future behaviour. On the importance of environmental factors also in relation to the identification of the most suitable treatment in terms of re-education.

corresponding animal species or it could also link descriptions of 'facts' (either natural or legal) present in possible court cases to the indication of corresponding decisions.

Indeed, the model in this case would not merely reproduce the examples present in the training set but would instead offer a generalisation of them: it could also be applied to new cases that differ in some respect from each of the examples on which the training was based.

Next, the model that is prepared by the learner algorithm is then used by a different predictor algorithm to provide hopefully correct answers to new cases.

If then the examples that are most similar to the new case (with respect to the features most likely to influence the outcome) have been answered, the predictor algorithm can propose the same answer in the new case⁷²⁰.

One can try to imagine how such tool could work if applied to the judicial sphere, in particular for predictive purposes. As the following paragraphs will show, if the cases most similar to the new case, with respect to the aspects most likely to influence the decision in one direction or another (depending on the model that is constructed by the system), have led to a certain outcome or decision, the same will then be proposed in the new case that presents certain similar characteristics and connections.

It would thus be a support to the judicial assessment, which naturally should then put in place the necessary individual reasoning, according to the general principle of combining statistical and probabilistic evaluation with the application of the rule in the light of the peculiarities of the case in question.

The proposal to introduce a sort of risk reduction treatment clearly looks to the possibility of applying such instruments to support the judge with the main idea of identifying the best sanctioning treatment that can best respond to the re-educative purpose of punishment, pursuant to Article 27, paragraph 3 of the Constitution. Clearly, this proposal, which remains on a merely theoretical level, leaves open and opens in itself questions that touch upon the fundamental guarantees and the procedural structure of the decision-making process. Therefore, in this last phase of development, one must ask oneself to what extent the penal system, which produces effects and affects personal freedom in a direct and strong manner, is prepared to rely, albeit in terms of mere support, on instruments governed by human beings but also by mathematical

⁷²⁰ "For example, if in the training set the animal images that are most similar (in classification-related aspects) to the new image proposed to the system are labelled as cat images, the new image will also be classified in the same way; if in the training set the loan applicants whose characteristics are close to those of the new applicant are classified as unreliable, the system will classify the new applicant in the same way; if in the past workers with characteristics most similar to the new applicant have been hired, the system will predict the hiring of the new applicant." thus, the example provided by G. SARTOR, *L'intelligenza artificiale e il diritto*, 49.

rules and procedures, the result of which, perhaps, cannot be fully contradicted by the evaluations of those who are then called upon to give the final answer on the prognosis.

The same decision, however, if left entirely to human beings, will continue to be intuitive and therefore potentially unreliable, but probably influenced by empathy and humanity, perhaps irreplaceable components in choices about personal freedom.

Now, it is precisely here that the importance of the twofold dimension of prognosis emerges even more clearly: on the one hand, it is not limited to assessing and quantifying the risk of reoffending, but also makes a decisive contribution to establishing how re-educational treatment should be oriented, starting with the choice of the type and extent of the sentence, to meet the actual criminogenic needs of the individual offender.

Clearly, the starting point for the proposition of a technological instrument within this type of assessment consists in a fundamental theoretical premise: that is, that every time it is necessary to predict the future behaviour of a given subject and, in particular, of the offender, it is unthinkable to proceed in an attempt to cancel out any margin of error; indeed, this type of prognostic judgements (as in reality any assessment made in criminal proceedings) possesses an inevitable probabilistic nature and therefore, a margin of uncertainty remains⁷²¹.

As already mentioned, it would at the same time be impossible to imagine the creation of a technological instrument (which may become better and better over the years) that could offer with certainty a solution for prognostic analyses and in particular for the assessment of a subject's dangerousness in relation to the choice of the best sanctioning treatment.

For this reason, what we actually want to propose here, albeit on a purely theoretical level, is to place alongside the data and factors that are considered most relevant and as free from discriminatory drifts as possible, those that pertain to the studies of psychology, psychiatry and criminology that can help lay the foundations for the algorithm for the formulation of prognostic judgements.

Underlying this is certainly the starting point that there is no theory capable of explaining any form of criminality except through a level of abstraction such as to lose any explanatory capacity with respect to the concrete case.

In fact, since the reliability of the prognosis does not depend exclusively on the theoretical framework of reference, but above all on the identification, connections and balancing of the individual factors selected and relevant in the case in question, one should try to identify the

⁷²¹ Indeed, as has recently been observed, 'available scientific findings on recidivism indicate that, to a certain level of accuracy, future behaviour can be predicted, and persistence in committing offences can be avoided'.

type of algorithm capable of detecting such instructions and data, and train it starting from a 'clean' data base and thus from a legal basis that has already 'corrected' the errors that have emerged.

As a continuation, it will be a matter of selecting (and here we will try to do so, albeit at a purely hypothetical level) those factors⁷²² that, on the one hand, are considered to be most 'detached' from subjective characteristics of the subject that are capable of invalidating the result and that, at the same time, can provide a sort of 'snapshot' that is at least capable of 'fixing' the situation at that particular moment.

As set out in the previous chapters, in the most recent studies on the use of artificial intelligence within the criminal justice system, the study of predictive algorithms at the service of the judge has mostly focused on the ability to calculate the risk of re-offending both in the commensuration of the sentence and for the application of precautionary or alternative measures. The use of algorithms is also extended to the choice of the type of rehabilitation programme in certain sectors such as that of sex offenders (which has been based in many jurisdictions, first of all in Canada, on the assessment of the risk of reoffending with systems based on Risk-Need-Responsivity - RNR, treatment must be proportional to the risk of committing a new offence), which use actuarial tools, based on empirically validated risk factors, drawn from the personal and criminal history of the subject (STATIC 99R, STABLE 2007, ACUTE)⁷²³.

2.1 *The output of the algorithm: predictions*

On closer inspection, the responses of a machine learning-based system are normally called predictions. However, the use of this term is not always correct and depends on the type of result that is generated by the algorithm. Indeed, in the context of machine learning, predictions

⁷²² K.S. DAHLE, *Psychologische Kriminalprognose*, Friburg, 2010, 14 ss.

⁷²³Through the RNR model, risk assessment tools are among the most common applications of AI technology to criminal justice, according to the 2018 Global Meeting on the Opportunities and Risks of AI and Robotics for Law Enforcement. These tools, which are already heavily used in Western 'correctional and probation services', on the one hand calculate, based on the identification and weighing of static (e.g. criminal history) and dynamic risk factors, the individual's risk or likelihood of reoffending, for crime in general and/or for specific types of crime, such as sex crime; on the other hand, this risk assessment is used to tailor 'treatment' to the need to modify the dynamic risk factors presented by the individual, or to respond to the criminogenic needs of the individual, in prison or on probation, as well as to influence probation decision-making and to monitor the individual after re-entry into the community. F. PEREIRA, *Artificial Intelligence, Offender Rehabilitation & Restorative Justice*, February 2020, in *The "Good" Algorithm? Artificial Intelligence: Ethics, Law, Health*. International Workshop organised by the Pontificia Academia Pro Vita, Date: 2020/02/26 - 2020/02/28, Location: New Hall of the Synod, Vatican City, lirias.kuleuven.be/2960856?limo=0.

do not always mean results or values that look into the future. In some cases, in fact, they are anticipations of the future, but in others the prediction itself concerns the present or the past⁷²⁴.

In addition, predictions in some other cases refer to an event that is likely to occur, regardless of the prediction itself; in other cases, it is a suggestion that may or may not be accepted by those who may realise the event itself.

Undoubtedly, another element that emerges is that a given system that makes automatic predictions operates in itself on the basis of correlations, that is, probabilistic relationships between input data and possible outcomes.

Indeed, a correlation consists in the fact that the presence of certain input data corresponds to a greater probability of a certain outcome. In such case, this is also referred to as a 'positive correlation' and at the same time, a lower probability of the same is referred to as a 'negative correlation'.

What is relevant is that such correlations are directly incorporated into the model built by the training algorithm, which, if it determines inputs that are positively correlated with a favourable outcome, it then tends to correspond a favourable prediction (the reverse is also true).

Normally, according to the general functioning of the algorithm, the effects of all relevant correlations that the system is able to consider are all aggregated together in a score, defined as a "score" that expresses the probability that in the case under examination the classification is positive or negative.

Indeed, what is noted by scholars of such systems is that in being able to assess the use of an automatic system that makes predictions, it is necessary to distinguish whether the data contained in the training set are constructed from past choices of human beings (for example, in the case of instruments used in public prosecution offices that serve to predict the possible commission of certain types of crimes in certain areas of the city) or from events independent of such choices⁷²⁵.

⁷²⁴ Think of a system that 'predicts' the classification of the content of an image or the authorship of a signature, thus on the point, G. ZARA – D. FARRINGTON, *Criminal recidivism: explanation prediction and prevention*, 5.

⁷²⁵ "Consider, for example, two systems used to evaluate loan applications. The first system has learned to evaluate such requests on the basis of a training set that associates information on past requests with the corresponding decisions by the relevant officials (acceptance or rejection). The second system, on the other hand, uses a training set that associates successful applications with the outcome of the loan (repayment or non-repayment). In the first case, the system learns to predict the decisions that the bank officials would have made in similar circumstances; in the second case, the system learns to predict the realisation of the desired outcome (the repayment of the loans granted). In the first case, the system reproduces the virtues (accuracy, impartiality, fairness) and vices (inaccuracies, biases, unfairness) of the officials; in the second case, it more objectively anticipates the desired or feared outcomes. This is the very illustrative example given by G. SARTOR, *L'intelligenza artificiale e il diritto*, 50.

It is precisely from the description of the proposed model that it emerges that the same evaluations and considerations could be made for a system that is intended to operate in the field of justice, for instance to determine whether or not to grant probation.

In fact, one could imagine, albeit at present only from a theoretical point of view, constructing an algorithm that could be trained either on the basis of a training set that associates the same probation requests with the corresponding decisions of the judges, or on the basis of a training set that associates the same probation requests with the subsequent behaviour of the defendant (thus indicating whether that person actually maintained a correct behaviour or whether he or she evaded punishment and thus repeated the offence).

Therefore, the assessment of the risk of recidivism and of the dangerousness of the subject must be supported using risk assessment tools that make it possible to combine statistical rigour and empirical experience.

Indeed, the role of the expert in risk assessment is accompanied by that of the judge to place, from time to time, this cognitive approach within the framework of the principles governing the penalty system.

The 'more probable than not' standard leaves a wide margin of discretion to the judge in predicting the future conduct of the accused/convicted person.

2.1.1 At which stage could predictions intervene?

On closer inspection, when one speaks of the inclusion of these instruments in the sentencing phase, one should probably envisage another type of sentence commensuration phase. In particular, one that refers to those theories based mostly on a sentence structure as a project.

As already anticipated, the predictive algorithm and, in particular, the risk assessment tool, could constitute a tool that, by providing a broader cognitive basis with regard to the state of the offender, would enrich the methods of responding to the offence in compliance with the fundamental principles of proportionality and the dignity of the person, realising the instances of individualisation of the sentence and giving substance to its re-educational function.

In this regard, it should be recalled that doctrine has for some time now been discussing the meaning of punishment as a project and, in particular, the precipitate of this theoretical perspective, namely prescriptive punishment. The discussion on prescriptive punishment, understood as a programme of intervention on the fracture produced by the crime and not as the infliction of evil corresponding to the culpable value, arises on the basis of those restorative justice orientations that want to build a path 'that promotes the offender's accountability with regard to the goods attacked and allows a reliable forecast of behaviour in conformity with the

law, by the same, for the future'. The aforementioned empowerment could be more effectively promoted, in the opinion of the relevant doctrine, by a prescriptive programme than by a prison sentence.

Prescriptive punishment, in particular, is characterised as a sanction with a project content, aimed at enabling the offender to critically review the fact committed. It is not a mere flexibilisation of the original prison sentence, but rather radically new means, constructed to prevent entry into prison, through the definition of a project for reacting to the offence in accordance with the objectives of tackling that specific criminal manifestation. In particular, the various obligations envisaged include participation in re-educational programmes, consisting of meetings with operators of the External Penal Enforcement Office, restitution, compensation and commitment aimed at eliminating the consequences of the offence, the provision of work of public utility and the carrying out of a therapeutic and socio-rehabilitative programme. Furthermore, of particular interest is the reference to the implementation of a programme, a sanctioning project, which may contemplate activities having a rehabilitative value with respect to the offence committed, in favour of the legal asset offended, of the person offended by the offence or the victims of similar offences; as well as participation in a criminal mediation proceeding with the person offended by the offence. In the face of a penalty with such a diverse content, algorithms could play an important role right from the moment the penalty is imposed. Incidentally, the method of inflicting prescriptive penalties seems to weld the demands of the substantive and procedural criminal law doctrine together, as both emphasise the importance of separating the phase of deciding responsibility from that of commensuration of the penalty, along the lines of what happens in systems that respectively know the separation between the moments of judgment in the strict sense and sentencing⁷²⁶.

With particular reference to the application procedure, the proposal formulated by the Working Group provides that, when the judge decides to apply a prescriptive penalty, he pronounces a sentence without determining the penalty, continuing in the same hearing or in subsequent hearings for the purpose of such determination. It is at this juncture, and this is the relevant point, that the defendant may formulate his/her own proposals on the content of the prescriptive penalty and may, above all, document his personal, family or social condition. The decision, in fact, comes after the judge has heard the defender, the assessments of the public prosecutor, any further discussions between the parties, and, if he so requests, the defendant.

⁷²⁶ In this sense, see the considerations of S. QUATTROCOLO, *Artificial intelligence*, 139. In the penalistic Italian doctrine, L. EUSEBI, *La pena tra necessità di strategie preventive e nuovi modelli*, in *Riv. It. Dir. Proc. Pen.*, 2021, 838.

Precisely in the light of the above considerations, it would seem possible that the judge would be able to perform his task more effectively if he could avail himself of ample and detailed information on the offender's risks and needs to frame the offender's condition and thus structure appropriate prescriptions. In this context, on closer inspection, a tool capable of providing a rich compendium of information on the offender could be the risk assessment algorithm.

If properly used, in fact, the risk assessment algorithm could be a tool to guide the judge in determining the sentence to be imposed, providing him/her with a spectrum of elements about the personality of the offender and his/her living conditions, so as to better define the contents of the prescriptive sentence, acting as a barrier to the 'noise'⁷²⁷ and 'biases'⁷²⁸ inevitably present in judicial decisions. Used in this sense, the AI tool would be able to become a real support for the judge, without replacing him/her. It could be the tool capable of filling the prescriptive penalty with content. Moreover, through a more precise characterisation of the situation to be managed, one could also respond to the flood of criticism directed to this new sanctioning model, accused of excessive vagueness in its content and therefore suspected of being entrusted to the absolute discretion of the judge in its determination, with a surreptitious emptying of the principle of the legality of punishment.

Clearly, the marginal considerations remain on the concerns that prescriptive sanction does not ensure adequately guaranteed limits of criminal intervention, in view of the areas of applicative discretion that it entrusts to the judge.

In conclusion, the content of prescriptive punishment rests on a clear premise: the commission of an offence reflects the existence of economic, social and cultural factors that contribute to the commission of the offence and, by virtue of this, it is understandable how a punitive treatment that is truly oriented towards the recovery of the offender and his resocialisation cannot disregard the need to affect these factors. The introduction of instruments of this kind constitutes, on closer inspection, a work of 'primary prevention', i.e. a strategy that intervenes first and foremost on the determining factors of crime and is the prerequisite for any criminal policy strategy that can be considered effective⁷²⁹.

⁷²⁷ Systemic 'noise' refers to random dispersion, all those errors inevitably present within a system or procedure that undermine its efficiency. On these issues, please refer to the work of D. KAHNEMAN - O. SIBONY - C. SUNSTEIN, *Rumore. Un difetto del ragionamento umano*, Milan, 2021.

⁷²⁸ On the particular profile of biases and fallacies that contaminate the judicial decision at the decisional stage, R. RUMIATI - C. BONA, *Dalla testimonianza alla sentenza. Il giudizio tra mente e cervello*, Bologna, 2018, 133 ss.

⁷²⁹ Concerning the Italian legal system in particular, it would seem evident that such results can only be achieved on the assumption of a real and effective cooperation between judges and experts⁶⁹, together with the provision of adequate resources both in the perspective of training technical operators and judges, and in the perspective of

As we have seen, it will be necessary to pay attention to the data that are the subject of the algorithmic basis, identifying the factors really connected to recidivism, asking the tool a clear question on the object of the forecast, giving centrality to training, to the training of the operators of the criminal justice system, in view of a broader collaboration in the resolution of problems that go beyond the limits of the individual disciplines and need to be addressed with an integrated approach.

3 *An initial critique of accuracy and the risk of generalisations*

It should be premised that the new technologies offer the advantage of being able to draw on and process huge amounts of data from sources such as jurisprudential and legislative databases, collections of precedents, and the like - with the use of sophisticated devices, which should make it possible to bring out relationships, coincidences, correlations, which allow to profile a person and predict his/her subsequent behaviour, even of criminal relevance.

Learning (in machine learning systems) is aimed at prediction, at the resolution of cases other than those analysed, which may occur in the future. This process is called generalisation: given a set of initial information, a rule must be extrapolated that is suitable for predicting and solving future cases that have not yet been analysed; machine learning aims at predicting a certain outcome⁷³⁰.

On closer inspection, from the first applications in the North American field, it was immediately possible to extrapolate the major problems associated with the use of such technological tools.

In fact, the risk of so-called implicit bias immediately emerged: on the one hand, where the input is not completely neutral, the output of the query runs the risk of being influenced by a bias, which may lead to the discrimination of individuals or social groups; on the other hand, the algorithm - which is conceived and interpreted by a human - may trivially reproduce unjustified social preconceptions. In fact, the greatest risk is precisely represented by the possibility of reinforcing the so-called implicit stereotypes, which are physiologically present in the person who has to make a judgement, thus increasing the risk of a criminal law approach of the type of author and enemy. If "sophisticated algorithms can anticipate the behaviour of certain subjects, [...] the risk is to fall into a new generation Lombrosian theory"⁴⁵, all in

stimulating greater investments in digitalisation, which seem feasible today following the recent reform proposals following the approval of the PNRR.

⁷³⁰ They report this definition in mathematical language M. BELKIN – D. HSU – S. MA – S. MANDAL, *Reconciling modern machine learning practice and the classical bias-variance trade off*, in *PNAS*, Vol. 116, 32, 2019, 158.

violation, first of all, of the principle of equality⁴⁶, of the principle of offensiveness of the criminal law of the fact (enshrined in art. 25, paragraph 2, Const.) and of the principle of guilt, correctly understood as guilt for the fact.

On closer inspection, it can be seen that the proposals of a system constructed to support the judge in choosing the best penalty treatment would be all the more individualised the more predictors it takes into account and, at the same time, the broader the training set.

Human judgements are also based on generalisations that are based on past examples, in which the judgment body has already been involved or of which it was aware. Hence, this is one of the reasons why support should be given since then absolute (correct and certain) individualisation risks and so is to escape even the human decision-maker.

Certainly, a further problem to consider concerns the fact of the empirical 'verifiability' of the result of the prognostic assessment, whether issued by a human being or with algorithmic support. There are, in fact, objective limits to the identification of reliable a priori probability data. Firstly, because this still remains a hypothetical and, secondly, because there is a lack of information that is necessary to correctly calculate the a priori probability of recidivism. The problem that arises, in fact, is that it is in no way possible to know the number of convicts with a positive prognosis of recidivism who, if they had remained at large, would have actually committed other offenses (and therefore realised the outcome of the prognosis)⁷³¹.

Thus, the impossibility of being able to verify the falsehood or truth of the positive prognosis of recidivism would risk increasing the risk that the judges, especially in all those cases in which they are uncertain about a given assessment, are then more inclined to order the restriction of the personal liberty, thus denying the benefit and applying the personal security measure.

Therefore, the positive prognosis of recidivism would be all the more reassuring for the judge in the concrete case and for the penal system in general because it would not be disproved.

On the contrary, however, errors in negative prognostic judgments, i.e. in cases where it is erroneously believed that the convicted person will not commit further offences in the future, risk then generating a 'loss of credibility of criminal justice in the community'⁷³².

⁷³¹ On the non-falsifiability of false positives, i.e. subjects whose behaviour is expected to be recidivist and who are therefore deprived of their liberty, without it then being possible to verify the correctness of this prediction, M. PELISSERO, *Pericolosità sociale e doppio binario*, 113

⁷³² *Ibidem*, 114.

Initial concerns have already emerged about the decisive significance of the algorithm in the decision-making process during sentencing. In fact, it has been observed how the algorithm is able to exert internal and external pressure on the judge who is about to formulate the decision, and how cognitive biases also encourage the use of such tools, considered objectively objective and neutral by virtue of the so-called automation bias.

Indeed, the most relevant issue concerns the fact that a judge, having to make a decision in conditions of uncertainty, will hardly allow himself to be conditioned by the outcome of the algorithm: in fact, already aware and 'influenced' by the high risk of recidivism, he will not run the risk of resorting to alternative means to imprisonment, let alone imposing too short a prison sentence⁷³³.

Finally, a further risk is that the judge may assess the subject as guilty not for the fact committed, but for the likelihood of committing offences in the future, by virtue of an argumentative inversion, or base his assessments on the personality profile of the defendant traced by the algorithm⁷³⁴, rather than on the facts actually committed and subject to trial.

Even more generally, there is also the fear that the use of AI tools contributes to a securitisation thrust of criminal law as an instrument of social control, considering people only as potential risks that can be controlled and constantly assessed (hence, as dangerous subjects).

⁷³³ Indeed, a judge who is provided with a risk assessment that predicts a high rate of recidivism 'might be led to impose a higher sentence without even the slightest awareness of the role played by "anchoring" in the decision itself'. Thus, on the point L. MALDONATO, *Algoritmi predittivi e discrezionalità del giudice: una nuova sfida per la giustizia penale*, 410.

⁷³⁴ On closer inspection, the doctrine (and also the Supreme Court of the District of Columbia) considers that the parameters set out in the Daubert judgment of the US Supreme Court should be adopted to establish the validity of a scientific law (in the application of Rule 702 of the US Federal Rules of Criminal Procedure), also in order to determine the validity, first of all, of the scientific theory of a psycho-criminological nature that should underlie the algorithm and then of the algorithm itself, as a translation of the theory into computational terms. In similar terms, authoritative Italian doctrine has suggested the use of the criteria of the Daubert judgment to enable the judge to assess the validity of a scientific law to be used in criminal proceedings, in particular for the purpose of ascertaining causality, considering that the Italian Supreme Court, while recognising in the well-known Bonetti judgment, on the Stava disaster, that scientific laws should "receive confirmation by recourse to rational and verifiable methods of proof", did not specify what these methods were; a suggestion that was in some ways followed by the subsequent jurisprudence of the Supreme Court. In order to guarantee then the accessibility and falsification of the 'algorithmic test', as well as the reliability of the predictive scientific method underlying the computational model, i.e. the correctness of the translation of the psycho-criminological theory (predictive scientific method) into algorithmic language, one must demand and ascertain upstream the validity of the scientific theory adopted in the calculation of dangerousness and the validity of the software model used. If one were to apply the criteria used in the Daubert judgment to affirm the validity of a scientific law in the light of the combination of the various criteria developed in the doctrine, one would have to ascertain (a) whether the expert's technique or theory has been (or can be) tested (i.e. whether the expert's theory can be empirically verified, or whether it is barely a subjective, apodictic approach that cannot be reasonably evaluated for reliability; the scientific methodological current "relies on the formulation of hypotheses on testing them to see whether they can be falsified"); (b) whether the technique or theory has been subject to publication and peer review; (c) the known (or potential) error rate in the application of the technique or theory; (d) the existence and maintenance of standards and controls; and (e) whether the technique or theory is accepted by the relevant scientific community.

The American criminal justice system⁷³⁵ aimed at incapacitating risk ends up perpetuating past discrimination (and, thus, perpetuating racial injustice). Such risks can be exacerbated using algorithms. And, again, the risks associated with the use of logics developed on a collective basis (with reference to groups or classes of people) that do not take into account the peculiarities of the single person emerge, also because to avoid false negatives such programmes are based precisely on a generic event ("programmers, ordinarily, choose as the final event object of the algorithmic prediction a generic event, such as to increase the basic dataset").

4 *Risk indicators: the difficulty of selection*

As already analysed when the concept of risk assessment was introduced, it was seen that risk indicators can be objective or subjective. Objective factors are not necessarily more reliable than subjective ones, for example, fear of women, an obviously subjective indicator, is one of those most strongly correlated in the literature with a high level of risk.

A first element and necessary step is to understand which are the relevant indicators, subjective and objective, that are considered useful to take into account. In particular, there are certainly certain types of offences (those which, in particular, present recurrent and repetitive elements over time) which offer, at first glance, a more 'complete' vision and immediately arouse greater fear. Certain elements, in fact, appear from the outset to be symptomatic of the risk of repetition of a certain behaviour.

It is believed that there are indicative elements that could, however, be readapted to different offences, but would only be more accurate for some of them (e.g. in cases where violence is a constant and recurring element). This is because they would characterise the perpetrated behaviour not as a mere and extemporaneous conduct by the subject on a given day and time, but, on the contrary, because they would be symptomatic of a risk that may possibly (more likely) occur.

For example, a selection of risk elements have been outlined by some authors⁷³⁶, which characterise a particular category of offences, i.e., those against violence against women. High

⁷³⁵ In fact, in the North American legal system, it is contested that such instruments are functional to the affirmation of the so-called New Penology, according to which the main purpose and responsibility of criminal law is 'the management of 'dangerous groups'', which replaces the so-called 'Old Penology', whose main purpose is the attribution of responsibility for specific criminal acts ('adjudication of guilt for specific criminal acts').

⁷³⁶ Among the risk assessments mostly used in domestic violence is the Spousal Assault Risk Assessment (SARA) by Anna Baldry. SARA is an instrument used on victims of domestic violence, developed in Canada by the British Columbia Institute on Family Violence and is used in 15 countries. SARA is defined as a 'guideline' that estimates the victim's level of risk (low, medium, high) and contributes to the planning of a victim safety programme. It

risk indicators⁷³⁷, for example, would include: fear of the woman criminal record, death threats, possession of or access to firearms, suicidal thoughts or attempted suicide, violence against children, increase in the frequency and severity of the episodes (escalation), expecting the relationship to continue forever, use of alcohol or substances, morbid jealousy, intention of the woman to separate.

On the other hand, among the non-specific indicators (with an average risk to be assessed overall on a case-by-case basis)⁷³⁸ have towards authority figures (real or perceived); inappropriate control and attention towards the partner; attribution of responsibility for violent behaviour to alcohol and substances; acceleration of involvement in the early stages of the relationship.

In fact, the type of risk factors must be taken into account in the assessment: static (such as addiction problems or criminal records) where the risk is rather constant over time, and dynamic (such as separation or employment problems) where the time factor can have a significant influence⁷³⁹. In order to determine the level of risk, it is necessary to assess not only the quantity of the factors present, but their interaction and development over time⁷⁴⁰. There are several specific risk assessment instruments that are based on the selection of individual and social, static and/or dynamic variables that are correlated with violence⁷⁴¹.

As partly already anticipated, one of the most reliable methods for risk assessment is the structured professional risk assessment based on guidelines and empirical studies on the subject: above all, the modifiable dynamic factors emerging from the scientific literature and the professional experience of the practitioner, deemed relevant for the specific case, are

consists of 20 risk indices, organised in five areas that detect both static and dynamic risk factors. SArA has demonstrated good validity and reliability, and it is also easy to use and thus accessible to various categories of operators. Overall, it is a flexible and sensitive instrument, which, however, suffers from the discretion of the operator administering it. There is also a short version of Screening (SARAS). Also, Increasing Self Awareness (ISA). It is a self-administered instrument for the timely detection of violence and sometimes also for assessing the risk of future aggression. It consists of a first calendar part in which violent incidents during the last year are noted down. The second part consists of a 20-item scale with dichotomous scoring, which assesses the victim's risk level.

⁷³⁷ On this point, J. C. CAMPBELL, *Risk assessment for intimate partner femicide. What practitioners need to know. Paper presented at the International, Conference on Children Exposed to Domestic Violence, London, Ontario, 2021.* A. N. WIESZ (et oths), *Assessing the risk of severe domestic violence. The importance of Survivors' predictions*, in *Journal of Interpersonal Violence*, 2000, 75-90.

⁷³⁸ J. C. CAMPBELL, *Risk assessment for intimate partner femicide. What practitioners need to know. Paper presented at the International, 2021.*

⁷³⁹ D. G. DUTTON – P. R. KROPP, *A Review of domestic violence risk instruments*, in *Sage Journals*, Vol. 1, Issue 2, 2000; J. Roehl – K. Guertin, *The current use of risk assessments in sentencing offenders*, in *The Justice systems journal*, Vol. 21, No. 2, 2000.

⁷⁴⁰ S. D. HART, *Evidence-Based Assessment of Risk for Sexual Violence*, 145.

⁷⁴¹ E. ALDARONDO – D. B. SUGARMAN, *Risk marker analysis of the cessation and persistence of wife assault*, in *Journal of consulting and clinical psychology*, 64(5), 1996, 1010-1019; N. Z. HILTON – G. T. HARRIS – M. E. RICE, *Predictive violence by serious wife assaulters*, in *Sage Journal*, Vol. 16, Issue 5, 2001.

analysed. It is therefore important to repeat the evaluation periodically, as circumstances, context and interventions change⁷⁴². It does not therefore express an evaluation in numerical or probabilistic terms but guarantees a more individualised framing of the case according to its peculiarities⁷⁴³. The tools that are used for this type of assessment are easy to use, take the form of checklists and are particularly effective for those working in contexts where an assessment needs to be carried out quickly and rigorous procedures are required (e.g. law enforcement agencies, health workers). They can be used in networking to structure a shared risk management project. Tools that respond to this method are for example: SArA, ISA and DA. The other method, which is valid in terms of reliability, is based on the use of actuarial tools that allow decisions to be made on the basis of a numerical score obtained on a predefined scale of static factors, which do not vary over time⁷⁴⁴. The scales identify the presence or absence of specific behaviours that provide a total percentile, referable to a precise level of risk, that can be compared with normative data. This method decreases human discretionary error, and the administration of the instrument is replicable at different times and by different assessors. These instruments are mainly suited for high risk and lethality levels, as they take less account of the dynamic factors of the specific case. The tools are well suited to the objectives of networking, in terms of structuring and sharing projects involving the application of severe protection measures. In order to choose the most suitable method and tool for one's purposes, one should keep in mind the primary goal of preventing future effects⁷⁴⁵, violent incidents, through a risk management strategy that includes a treatment phase, monitoring and constant supervision⁷⁴⁶. An important purpose in using risk assessment tools is to strengthen the collaboration and transparency between the different services involved in the care of violence situations⁷⁴⁷, as methods and tools can be used by different professionals, including nurses, psychologists, social service workers, anti-violence centres and law enforcement. It is therefore important that these tools are able to capture information that is useful for different purposes: securing the victim, recommendations for the detention or release of the defendant, treatment of the aggressor.

⁷⁴² C. CAMPBELL, *Risk assessment for intimate partner femicide. What practitioners need to know. Paper presented at the International.*

⁷⁴³ *Ibidem.*

⁷⁴⁴ J. BONTA - M. LAW - K. HANSON, *The prediction of criminal and violent recidivism among mentally disordered offenders: A meta analysis. Psychological Bulletin*, 123, 123–142, 1998; W.M. GROVE - D.H. ZALD - B. S. LEBOW - B. E. SNITZ, - C. NELSON, *Clinical versus mechanical prediction: A meta-analysis. Psychological Assessment*, 12, 19–30, 2000.

⁷⁴⁵ secondo Hart (2010)

⁷⁴⁶ P. R. KROPP, *Intimate partner violence risk assessment and management. Violence and Victims*, 2008, 23, 202.

⁷⁴⁷ S. D. HART, *Evidence-Based Assessment of Risk for Sexual Violence*, 145.

4.1 *The most sensitive issue: a chance to overcome. The choice of risk factors*

On closer inspection, as analysed in the preceding paragraphs focusing on the analysis of prognostic evaluations, any prognostic judgement to be made presupposes the collection of all the factual elements, the characteristics of the offender and any other data on the basis of which it is then possible to establish which predictive factors of recidivism exist in the concrete case.

Indeed, the correctness and explicative capacity of the prognostic judgement on the likelihood of reoffending and on the special-preventive effects of the sanctioning response depends on the comprehensiveness and relevance of these elements that are selected. Clearly, parallel to a correct and complete collection of data, there must be a complete nomological knowledge, based on laws and general principles formulated on the basis of consolidated knowledge that are then able to attribute a certain meaning to the collected data and to establish the impact of these factors on the offender's future behaviour.

Undoubtedly, the first step of gathering all the necessary information on the predictive factors of reoffending appears both feasible and achievable. Indeed, this objective must be placed within the framework of criminal proceedings which, besides being governed by the requirements of reasonable duration and procedural economy, is at the same time subject to the principle of the presumption of innocence set out in Article 27(2) of the Constitution.

This premise consequently implies that the ascertainment of predictive factors entails the entry of complex assessments that are entrusted to experts in extra-legal knowledge, with an obvious increase in time and costs. Risk factors, as already analysed in the previous pages, can allow an actuarial (or statistical) approach to the assessment of criminal dangerousness⁷⁴⁸. In fact, through a combination of them, one can set up 'scales' that allow the attribution of a score to the subject under examination⁷⁴⁹.

As already mentioned, one of the most delicate tasks in the type of prognostic assessment is precisely that concerning the identification and characteristics of the offender and the situational factors that led to the commission of the offence and that are likely to cause a relapse.

⁷⁴⁸ On this point, A. M. MAUGERI, *L'uso di algoritmi predittivi per accertare la pericolosità sociale*, 12.

⁷⁴⁹ The 'scales', which are used for the actuarial assessment of criminal dangerousness, differ from each other, depending on the population in relation to which they have been developed⁴⁰; the type of offences involved; the timing of the risk (immediate, or medium- or long-term); and the application context. Risk factors, in turn, can be static, which cannot be modified (e.g. gender and ethnic origin); stable dynamic, which are modifiable through therapeutic treatment (e.g. impulse control); acute, which change rapidly and are associated with a condition facilitating the violent reaction (e.g. drug use). In fact, there are generic scales, i.e., relating to all offences, and specific scales, relating to individual types of offences, such as sexual offences or violent offences. See on this point, F. BASILE, *Intelligenza artificiale e diritto penale*, 17; S. QUATTROCCOLO, *Artificial intelligence*, 148 ss.

Reference is made, in particular, to the predictive factors on the basis of which the prognostic assessment should then be made. These are thus factors which, in the light of the empirical studies carried out today, offer consolidated essential information on the probability of recurrence⁷⁵⁰.

Among these, there is a core of four factors, which are also referred to as 'the big four' and which are considered to be the most decisive:

1) The previous antisocial or criminal behaviour, the so-called history of antisocial behaviour which includes early involvement in various and numerous social activities

2) The antisocial personality pattern which includes personal characteristics such as aggressiveness, impulsivity and lack of self-control

3) The antisocial attitudes, values and behaviour, the so-called antisocial cognition, i.e. identification with criminal patterns or rationalisation of a wide range of circumstances in the presence of which the offence was committed

4) The frequentation of antisocial environments, the so-called antisocial associates, which also means realising a relative isolation of non-criminal environments⁷⁵¹.

In addition to the main factors that were considered to be the most relevant, there are four others that are, however, generally less relevant for the future prognosis of reoffending. These include the family environment (family/marital circumstances), poor performance and involvement with consequent lack of satisfaction at work and school (school/work), low level of involvement and satisfaction in non-criminal leisure activities (leisure/recreation), and alcohol or drug abuse (substance abuse).

The next objective, once the predictive factors have been identified⁷⁵², is to identify coverage laws, derived from generalised data of experience that can guide the prognostic judgement. Indeed, such coverage laws, which are formulated based on the observation of a large number of similar cases, can partly 'fill with meaning' the extremely vague general clauses formulated by the legislature, when the judge is asked to predict the future behaviour of the offender.

⁷⁵⁰ One of the most comprehensive works on this point is that of D. A. ANDREWS – J. BONTA, *Rehabilitating criminal justice policy and practice*, in *Psychology, Public policy and Law*, 16(1) 2010, 39-55. Among the many further researches dedicated to specific categories of offenders see P. GENDRAU - T. LITTLE - C. GOGGIN, *A meta-analysis of the predictors of adult offender recidivism: What works!*, in *Criminology*, 1996, 575 ss.; J. BONTA - M. LAW - K. HANSON, *The prediction of criminal and violent recidivism among mentally disordered offenders*, 123; R.K. HANSON - M.T. BUSSIERE, *Predicting relapse: A meta-analysis of sexual offender recidivism studies*, in *Journal of Clinical and Consulting Psychology*, no. 66, 348 ss.

⁷⁵¹ D. A. ANDREWS – J. BONTA, *Rehabilitating criminal justice policy and practice*, 58 ss.

⁷⁵² For a more complete analysis on the type of factors and, in particular, on the differences and characteristics of static and dynamic factors, see Chapter III.

Indeed, the most complex part is precisely that of assessing the prognostic impact of the risk factors on the offender.

This is because on the one hand individual differences interact in a peculiar way with those particular facts, but also because they accumulate and vary over time according to the developmental stages of the subject and then because their effect is inevitably influenced by the external environment and situation.

5 The forward-looking perspective and its structure: individualised judgement and human control of the judge

In order to overcome the aforementioned risks of discrimination and generalisation, in which the requirements of a correct individualised judgement would be undermined (essential for ascertaining the social dangerousness in criminal matters), the use of a two-phase system should be considered, which, in addition to the dataset relating to the criminal profiles deemed "similar" to that of the defendant, also takes into account the answers provided by the latter during the interrogation (whereas this is not the case, for example, with the PSA, pretrial risk assessment), or in any case of a series of factual elements relating to the case and the subject in question. "It is a question of enhancing the 'individualising' moment in order to ferret out the doubts of a possible violation of the right to an individualised sentence"⁷⁵³.

In order for the decision taken in the light of the results of an algorithm to comply with the principles of due process, human supervision of the algorithmic decision is always necessary, as also stated in the document prepared by the Expert Group appointed by the European Commission. Human oversight is a necessary safeguard to guarantee the reliability of the AI and should make it possible to overcome some limits encountered by the transparency of the algorithm's functioning (such as the secrecy protected by intellectual property rights, which in the judicial field should however be dispensed with, but also the need to avoid making the system flawed and potentially subject to manipulation or hacking attempts), bearing in mind

⁷⁵³ The Loomis case also shows that the judge could not have based his decision solely on COMPAS, because of "the need for the adjudicating body to apply the results of the programme by exercising its discretion on the basis of balancing against other factors". In that case, the U.S. court held that there was no violation of due process precisely because the assessment made using the A.I. system constituted only one element of the judgment, not decisive and corroborated by other factors, also because "the comparative-statistical nature of the software used was not capable of guaranteeing a completely personalised judgment, but only the result of the analysis of statistically similar circumstances and experiences". See OCCHIUZZI B., *Algoritmi predittivi: alcune premesse metodologiche*, in *Riv. Trim-Dir. Pen. Cont.*, no. 2/2019, 397 ss.; In fact, the judgment states 'COMPAS's assessment of the risk of reoffending does not express the specific probability that an individual offender will reoffend. On the contrary, it provides a prediction based on a comparison between a set of data on the subject and a set of similar information' (State v. Loomis, 881 N.W.2d par. 15).

that "the degree of explicability required largely depends on the context and the severity of the consequences in case the result is incorrect or otherwise inaccurate. For example, if the purchase recommendations generated by an AI system are inaccurate, this does not raise major ethical concerns, while the situation is different when AI systems have to assess whether or not to grant parole to a person sentenced to imprisonment".

In fact, pursuant to Article 111(6) of the Italian Constitution, the critical control instrument of the exercise of binding discretion is the statement of reasons, at the centre of which are the reasonableness and validity of the arguments according to an itinerary of rationality that must always be supported by a "verifiable hermeneutic foundation". The obligation to state reasons represents a legal guarantee, an imperative addressed to the judges, which fulfils a dual function of control over the decision of the judicial body, both intra- and extra-trial. The fundamental intra-trial control is linked to the possibility for the parties to challenge what the judge has decided in the various stages of the process and is linked to the right of motivation. In the Italian legal system, however, the activation of ex officio investigative powers, or even at the request of a party, is precluded by the provisions of Article 220, paragraph 2, of the code of procedure, which provides that "expert opinions are not allowed to establish the habituality or professionalism of the offence, the tendency to commit offences, the character and personality of the defendant and in general the psychological qualities independent of pathological causes"; considering that the use of an algorithm would correspond to an expert opinion, it would be unusable for violation of the law pursuant to Art. 191(1) of the Code of Criminal Procedure. Indeed, Article 220 of the Code of Criminal Procedure indicates some of the elements that the judge should take into consideration pursuant to Article 133(2) of the Code of Criminal Procedure, Article 220 of the Code of Criminal Procedure in fact indicates some of the elements that the judge should take into consideration pursuant to the second paragraph of Article 133 of the Criminal Code, excluding the possibility of carrying out technical-scientific investigations.

It follows that, under these provisions, any output produced by the AI can only be considered as a mere clue, which must always be corroborated by other elements of evidence, and that any use of the algorithm in court must be subject to significant human control. It being understood, however, that the same Article 220, c. 2 does not prejudice "what is foreseen for the purpose of executing the sentence or the security measure", opening to the use of expertise and therefore of the algorithm, if the use of the algorithm is equated to the expertise, as well as "of psychology, criminology, and related disciplines" in the provision on security measures and in the executive for the choice of alternative measures or the type of treatment most suitable for

the particular needs of the convict on the basis of the "scientific observation of the personality" (Art. 1 and 13, Law No 354/'75 and art. 28 Presidential Decree No 431/'76) (and opening, according to a certain approach, to the biphasic process). All this while ensuring compliance with the guarantees of the criminal trial, starting, as mentioned, with the right to cross-examination (the defence must be in a position to know the weight attributed to the various factors and to verify the reasonableness of the parameters and the method used, as well as the correctness of the final assessment) and the control of the judge on the basis of the principle of free conviction, the judge who should motivate and account for the choice of the data entered in the system in the light of the concrete fact.

On the other hand, however, the sector studies would seem to show how the actuarial assessment of the offender's risk of reoffending is much more accurate than the human one, as it is able to process an immense amount of data of which no judge could reasonably dispose⁷⁵⁴. It is then a question of assessing whether the use of such algorithms may ultimately prove useful in ensuring greater reliability of social dangerousness judgments or whether the risks outweigh the possible benefits, remembering that 'the core of any risk assessment software is the psycho-criminological theory that inspires it [...] it is up to the software designers to devise the best model to provide a rapid and reliable response regarding dangerousness, and above all, that in its essence digital risk assessment can be compared to expert testimony: the software, however, allows for more questioning, much more quickly, without the need to rely on the personal presence of an expert'. 'The fundamental question is, therefore, whether risk assessment software should be treated as psycho-criminological expertise'.

In conclusion, the critical issues that have emerged in the US system demonstrate the importance of full *disclosure* of the evaluation indices and weighting factors used. In the post-conviction phase, the defence must be in a position to know the weight given to the different factors and to verify the reasonableness of the parameters used, as well as the correctness of the final evaluation. The technical assessment should be performed in compliance with the principle of cross-examination (Article 111(2) of the Constitution). It is no coincidence that in the proposal for a regulation in the European context, Article 14 focuses attention on the fundamental problem of human control, highlighting the need for the person called upon to adopt a decision based on the output of the high-risk AI system to be aware of the limits

⁷⁵⁴ C. BURCHARD, *L'intelligenza artificiale come fine del diritto penale? Sulla trasformazione algoritmica della società*, 1933 observes that certain 'doubts about discriminatory use concern only the concrete implementation, but not the fundamental normative-regulatory concepts of an effective and efficient algorithmic protection of legal goods as well as of an objective, neutral and consistent application of criminal law'.

inherent of the technology employed and even to be able to decide against such outcomes if the concrete case requires it. The judge, therefore, in his capacity as *peritus peritorum*, will thus be able to carry out a check on the use of methods with a certain reliability, taking a position on the objections raised by the parties. However, it cannot be overlooked that the implementation of such a model depends, to a large extent, on the existence of valid risk assessment tools. Only the convergence of interdisciplinary expertise between psychological and criminological diagnostics, actuarial science and technology will allow the creation of a tool capable of rationally balancing the results of personal investigation with sufficiently large data samples.

5.1 *A key to begin with: the algorithm applied only in bonam partem*

Given the difficulties and major obstacles that are to be overcome at the outset, it is considered extremely useful and necessary to initiate and follow up on such a reflection, so as to enable the analysis to be focused in broad terms and on a research that certainly needs a much longer path.

Notwithstanding this, one can see how at the margins and in a merely embryonic phase of such research emerges the possibility that one wishes to reiterate and support here, which can also be accompanied by a possible application of such instruments circumscribed not only in a particular phase of the proceedings, but also to a restricted application only for purposes in *bonam partem*.

Indeed, considering that the criminal justice system can benefit from the contribution provided by other sciences and from the use of algorithms, but, at the same time, must preserve its constitutional structure, it is considered necessary to verify with great caution and with respect of the principle of proportion if the use of Artificial Intelligence systems can satisfy not only the need to increase the productivity and efficiency of the judicial system, and in particular the reliability of assessments of dangerousness, but also the protection of citizens' rights. This would require proper and careful regulation to ensure that the use of technology remains at the service of humanity⁷⁵⁵. In this direction, for example, limits should be established on the possible use of algorithms in the sense that, without prejudice to the fact that, as things stand, their use in the Italian criminal justice system could only be possible in the aforementioned terms when assessing dangerousness, if, *de iure condendo*, its use should also be admitted in the commensuration of punishment, its use should be limited only *in bonam partem*. In fact,

⁷⁵⁵ F. T. RIZZI – A. PERA, *Balancing Tests As A Tool To Regulate Artificial Intelligence In The Field Of Criminal Law*, in *Special Collection and Artificial Intelligence*, 2021.

although the penalty must be conceived by the legislator and commensurate by the judge (in the choice of type and measure) in order to pursue the re-educative function pursuant to Article 27, c. 3 of the Italian Constitution, in its concrete commensuration it finds its limit in the degree of culpability expressed in the fact committed. Respect for the principle of culpability pursuant to Article 27 of the Constitution allows the intervention of special prevention requirements only for the purpose of a possible mitigation of the punitive intervention; it will not be possible to aggravate the sentence in consideration of an accentuated capacity to commit offences pursuant to Article 133, c. 2145, and, therefore, of the risk of recidivism.

The possible use of algorithms, then, could be limited only *in bonam partem* to possibly allow mitigation of the sentence in the presence of a limited capacity to commit offences, in compliance with the principle of personal criminal responsibility under Article 27 of the Constitution; 'risk assessment results are used to identify treatment needs or to mitigate sentences, but not to enhance sentences'. The use of algorithms, then, could also serve to identify the most suitable treatment in view of the characteristics of the subject. The doctrine, in fact, already advocates the use of actuarial risk assessment systems, such as Statist 99R, Stable 2007 and Acute, for the selection of the most suitable treatment for the level of risk and which is truly adequate for the criminogenic needs of the participants, in the light of an assessment conducted according to an evidence-based model, such as the cognitive-behavioural model (Risk-Need-Responsivity).

To give concreteness to the re-educative purpose of punishment and with a view to overcome also the discriminatory effects of prognoses linked to the past behaviour of the subject (in the logic of the criminal law of the enemy or, of the so-called New Penology as already mentioned), - whether they are of an intuitive nature or are made by making use of predictive algorithms -, one could, then, conceive a different approach in the management of risk and, therefore, of dangerousness, which is not merely control based on detention or surveillance measures. In the logic of the principle of proportionality and extrema ratio of penal intervention, it is a matter of adopting a solidaristic approach proper to a welfare state, which has the obligation to implement the principle of substantive equality pursuant to Article 3 of the Constitution, such as the so-called Supportive Response to Risk, i.e. the offer of support tools to eliminate risk factors. This perspective is already pursued by actuarial risk assessment systems, such as Statist 99R, Stable 2007 and Acute, for the choice of the most suitable treatment for the criminogenic needs of participants. Most recently, the Good Lives Model (GLM) 149 takes an even more 'supportive' approach, "a strengths-based rehabilitation theory

that aims to equip clients with internal and external resources to live a good or better life—a life that is socially acceptable and personally meaningful”.

Such a model has been particularly tested in the treatment of dangerousness of sex offenders, provided that the subject is willing to change. This model starts in general from the recognition that "risk as the product of social conditions should lead us to seek responses that directly address those conditions." In the light of a socio-structural analysis of the risk, in this logic also the use of predictive algorithms for the evaluation of the dangerousness and the choice of the best treatment will be functional in not increasing a model of a police and repressive state, but a model of social state of law because it will lead to the implementation not of restrictive measures and social control, but measures aimed at recovery and social reintegration.

6 Comparing rights: the feasibility of a proposal between a balance of rights and guarantees

In expressing an opinion in favour of the introduction of statistical-actuarial assessment tools for the commensuration of sentences, it seems necessary to emphasise the need to ensure respect for the defendant's defence guarantees.

Indeed, the use of these systems and the probable and future proposal of such tools within the justice system is confronted with innumerable issues concerning, first of all, the guarantees and institutions of criminal law and, secondly, the protection of the personal data of individuals, since these tools use and process huge amounts of data. Processing that must in any case be subject to the rules laid down at national and supranational level for the processing of personal data and at the same time detect the greatest points of friction with them.

An attempt has been made to explain, especially in the first part of this paper, the reasons behind the decision to focus the study on the assessment of dangerousness and how it also indirectly affects the choice of the best punitive treatment for an individual. In fact, the judgment of social dangerousness for the purpose of applying security measures presents the serious limits of scientific reliability of criminogenic and predictive judgments, to the point that, it is observed, 'it can provide pseudoscientific bases for those forms of discrimination - social, political, cultural, religious, racial - so frequent in the twilight of democracies'; so much so that it is not surprising that the criminal law of recent years, modulating itself on the telos of security, a meta-objective projection of danger, has often identified its elective targets precisely in the weakest: in migrants, the marginalised, the homeless, street prostitutes, graffiti artists, drug addicts. The concept of dangerousness, in fact, is configured as 'hybrid', connoted simultaneously by medical and legal parameters, very ambiguous and scientifically

inconsistent. The Supreme Court itself is aware of the fact that "the assessment of social dangerousness (Article 203 of the Criminal Code, which recalls, in paragraph 2, the inescapable declination made by the legislator in Article 133 of the Criminal Code) is, like any prognostic judgement, based on the appreciation of the recurrence of a "danger", which is by its nature a judgement aimed at the future, which excludes its possible declination in terms of historical certainty (an attribute with which one can, conventionally and procedurally, qualify only past conduct), with an ineradicable margin of fallibility."

In reasoning towards this direction, it has been seen that the positive aspect of algorithms is that they 'design a normative procedure that moves from a set of data towards a desired output, excluding subjective intuitions and arbitrariness from the process'. In this way, it represents a mathematical model, which can be run by a human being, even in a criminal case, provided that it is based on a validated theory and that this theory has been correctly encoded in the algorithm. These two requirements are crucial: [...], the possibility to review, discuss, challenge algorithms is a basic condition for fair criminal proceedings, in accordance with fundamental human rights.

Indeed, the reasoning initiated in the North American field has very precise historical origins and moved towards an attempt to overcome the so-called sentencing malpractice.

Therefore, the criteria developed move from the need to balance the protection of industrial secrecy with the principle of transparency, to the guarantees of verifiability (auditability) and consistency (consistency) of the data used and the outputs produced. Under the aspect of the verifiability of the algorithmic decision - an objective that in the Italian penal system could be attainable with the ostension of the motivation in the jurisdictional measure, a burden that should be "strengthened" in the case of the choice of the hermeneutic option most unfavourable to the defendant. In the US doctrine the focus is placed on solutions that allow the result provided by the algorithm to be "crystallised" and to collect and preserve this data to protect the right of defence of the defendant who will thus be able to access it at a later date.

In particular, with reference precisely to the reasoning that the judge is required to sign and, specifically, on the decision on the quantum of penalty to be imposed in the concrete case, the magistrate will be required to give in his reasoning the results of the assessment and the reasons why he/she believes he/she must agree or disagree with its results. The model of conditional discretion allows more objective elements of assessment, logically verifiable and censurable by the parties with the ordinary means of appeal, to enter the trial dialectic. The magistrate's discretionary scrutiny eliminates the risk of non-individualised treatment due to the use of statistical findings. In fact, he/she will be required to critically assess both the reliability of the

risk score and the objections raised by the defence counsel and the public prosecutor, with the possibility of departing from the results of the assessment or taking them into account only in part for the determination of the penalty to be applied to the concrete case. A system structured in this way would make it possible to 'absorb' into the heart of the trial dialectic and the motivation of the conviction the questions and doubts relating to the instrument used to assess the offender's capacity to commit offences.

There is, however, the arduous choice on the identification of "who" will have to carry out this task and "how" the information will be collected, joints that will be crucial for an effective "understanding" of the algorithmic decision by those who will have to challenge it. The principle of the consistency of the decision then aims to guarantee the identity of the output over time: the objective is to guarantee that the system repeating the operation at a later time using the same inputs will always generate the same output, although in the meantime the algorithm has evolved thanks to machine learning. It is precisely the algorithm's self-learning capacity, in fact, that determines the risk that the same dataset used several times by the risk calculation tool will generate a variety of dissimilar risk scores. The diversity of the evaluations generated over time, even if always based on the same inputs, would thus depend solely on 'chance', i.e. the moment in which the calculation is made. Transposed into the Italian penal system, an AI algorithm for calculating dangerousness or recidivism would also clearly conflict with the principle of equality under Article 3 of the Constitution.

In this concluding chapter, an attempt will be made to summarise the *de iure condendo* proposal to be envisaged and, at the same time, the regulatory panorama with which A.I. instruments must be confronted today.

In this regard, an attempt was first made to identify the proposed application model (although necessarily limiting its analysis) to a descriptive and more summary analysis, to identify its application moment and essential function. In a second moment, in the part following the descriptive analysis, the attention will focus on the boundaries drawn from the criminal law, its institutions of guarantee and the procedural norms. In conclusion, an attempt will be made to analyse all those issues that are connected to the use of personal data and therefore to all the regulations relating to it.

The aim of this chapter is therefore to show that what has been illustrated and proposed in the previous chapters is possibly achievable by remaining within the legal framework that currently exists. Furthermore, these concluding reflections emphasise the urgent need for reflection on this topic.

Clearly, the choice to deal in this study with the possibility and at the same time possible implementation of a computational modelling of A.I. in Criminal law leads to the attempt to provide a clear theoretical and conceptual framework for reflection.

For this reason, in fact, it is first necessary to assess the national legal landscape and in a second step, the European position regarding the possibility of actually introducing such tools in criminal law systems⁷⁵⁶.

Indeed, the possibility of selecting the possible options in the choice of sanction in order to identify the instrument that could best reduce the possibility of reoffending through re-education, or incapacitation, or deterrence, is a scientific question that should be informed by the science of best practices, i.e. 'evidence-based practice', defined as: 'professional practices that are supported by the "best research evidence", consisting of scientific results concerning intervention strategies [...] derived from clinically relevant research [...] based on systematic reviews, reasonable effect sizes, statistical and clinical significance, and a body of supporting evidence. Thus, the concept of evidence-based practice in corrections refers to corrections practices that have been proven through scientific corrections research 'to work,' to reduce offender recidivism ".

It is observed, again, that evidence-based decision-making is part of a broader pattern in contemporary society involving the use of scientific research to improve the quality of decision-making⁷⁵⁷.

6.1 The narrower frame of applicability: the possibility of intervention in the face of relevant factors with dynamic characteristics

From the analysis developed thus far, the primary insuperable difficulty of identifying the factors to be taken into account certainly emerges; once this step has been overcome, however, one comes up against another element and another stumbling block that points to one of the main research questions.

The question revolves around whether or not, once the algorithm and the main risk factors to be taken into account have been identified, it will be possible to envisage a 'universal' application of such algorithms, in the sense of being able to apply them to all types of offences and in particular offenders.

⁷⁵⁶ On this point, S. QUATTROCOLO, *Artificial intelligence*, 17 s.

⁷⁵⁷ R. K. WARREN, *Evidence-based practices to reduce recidivism: implications for state judiciaries*, in *Crime & Justice Institute* 2007, 20; R. E. REDDING, *Evidence-Based Sentencing: The Science of Sentencing Policy and Practice*, in *Legal Studies Research Paper Series*, Paper No. 09-41, 2.

The question comes to the fore because the very subject of the judgement is changed, which then shifts from prognosis of future behaviour to ascertaining the characteristics of the offender and situational dangerousness.

Indeed, in order to make the prognostic judgement it does not then seem sufficient to ascertain with certainty the characteristics of the personality, and then to consider that the situational factors have a simple possibility of occurrence, but it is then necessary to identify the characteristics that, with a view to the future, are criminogenically relevant for the purpose of identifying the risk of reoffending.

Indeed, for example, one could imagine an offender who has a long criminal record and a serious lack of self-control, a so-called 'pro-criminal' attitude and is also a drug addict.

In this case, if the offences committed are linked in some way to the state of drug dependence, for which the subject is undergoing a therapeutic programme, the other personal characteristics lose their significance in the context of the prognostic judgement.

What is therefore relevant is the intervention on the dynamic predictive factor that is represented by the state of drug dependence. Therefore, in this sense, certainty about personal characteristics would not offer information about the standard of ascertaining prognosis. Instead, it would offer information about the present that is not necessarily relevant to the assessment of future behaviour.

Therefore, a distinction would have to be made between, on the one hand, the signs, i.e. those relevant risk factors that would have to be taken into account and, on the other hand, the evaluative biases, i.e. those elements that can and do lead to erroneous conclusions.

7 *The main issues arising from the first reflections*

On closer inspection, from the initial methodological premise, the need to channel and identify the major friction points of these instruments with the rights at stake emerges. Undoubtedly, the question on which data are analysed and which tools the subject possesses to know the data also (and above all) has been judged.

Therefore, on a critical level, questions arise relating to the limits of constitutional guarantees, fundamental rights that also receive protection from other normative sources; the question on the transparency of data, the selection of data and at the same time the elimination of so-called 'dirty data', which then entail as the first consequence the invalidation of the result; the neutrality of the algorithm.

7.1 *Constitutional limits*

The dangerousness of a subject is deduced exclusively from the behavioural patterns and decisions taken in a given community in the past, in contrast with the principle of individualisation of the punitive treatment, pursuant to Article 27, paras. 1 and 3, of the Constitution, as well as, of the canon of individualisation of the precautionary treatment, derivable from Articles 13 and 27, para. 2, of the Constitution.

The consideration that the inclusion in the algorithm of information relating to previous convictions, sentences already served, ends up disregarding the re-educative function of punishment, for example when bail is denied on the basis of the risk assessment calculated by an algorithm also considering such previous convictions.

In the first place, the consideration that in the Italian penal system the commensuration of punishment is entrusted to the judge's discretionary assessment is undoubtedly relevant. Indeed, entrusting the prognostic judgement in terms of the seriousness of the offence and the capacity to commit offences (pursuant to Article 133 of the Criminal Code), as already mentioned, would risk colliding with certain fundamental principles of the system.

In fact, as already noted when analysing the issue from a comparative perspective, compared with the American penal system, in the Italian legal system there is no temporal caesura between the pronouncement of the sentence and the imposition of the penalty (nor even a phase of 'investigation' into the personality of the offender)⁷⁵⁸.

Therefore, and it is here that a point of collision is to be found, the idea that an algorithm can replace or even just support the judge in assessing an offender's risk is already limited and precluded by several constitutional limits. In particular, in addition to those statutes already found in Art. 25, 102 Const., they are also found in Art. 101, para. 1 Const. which states that judges are subject only to the law. This means that the judge cannot be bound by the outcome of algorithmic procedures⁷⁵⁹ if one sticks to the literal tenor of the rule.

Another limitation of such hypothetical introduction is to be found in Article 111(4) of the Constitution, which guarantees cross-examination in the formation of evidence, thus preventing the judge from acquiring or evaluating elements other than those subject to cross-examination by the parties.

⁷⁵⁸ See L. D'AGOSTINO, *Gli algoritmi predittivi per la commisurazione della pena*, in *DPC*, No. 2, 2019, 267 e ss.

⁷⁵⁹ F. DONATI, *Intelligenza artificiale e giustizia in Rivista Associazione italiana dei Costituzionalisti*, in *Rivista AIC*, No. 1, 2020, 428.

7.2 *Compression of personal liberty: between rights and balances under Article 5 ECHR*

On closer inspection, the introduction of such instruments in decisional phases that are not only very delicate but which, above all, affect the personal freedom of a subject, could undoubtedly and do indeed affect the right to personal freedom, which finds its cornerstone of protection in Article 13 of the Constitution and Article 5 of the ECHR.

As already mentioned, the pervasive force that such tools may have affects not only the risk that they may touch the individual's personal data and that a violation may ensue, but also that such decisions may affect the individual's personal freedom, this must therefore be balanced by the guarantees that come to rebalance that part of the right which is even only partially affected.

The guarantees laid down in Article 5 ECHR⁷⁶⁰ come into play: in particular, that of access to the judge and the obligation to state reasons⁷⁶¹.

As already mentioned, the guarantee of access to the judge, which branches off along two lines, is fundamental in this regard⁷⁶². As already pointed out⁷⁶³, if a new subject is introduced into the trial or an artificial intelligence instrument, this same form of guarantee must be declined in an initial contact that the subject must necessarily have with the judging body. In fact, one cannot imagine a decision affecting the personal freedom of the subject where the subject does not decide after having been able to contact the magistrate. On the other hand, this guarantee includes all the other complementary guarantees, such as, for example, the setting of a hearing in which the detainee, assisted by his defence counsel, must be able to participate and, in turn, must be able to assume full knowledge of the reasons why a particular decision was taken, through the instrument of the statement of reasons⁷⁶⁴.

A second form of access to the courts is the right to an effective remedy before a court to assess the legality of the measure. This is, after all, a procedure that implies the application of

⁷⁶⁰ In particular, reference is made to Paragraphs 3 and 4 of the provision, which in turn provide: Paragraph 3. Every person arrested or detained in accordance with the conditions laid down in paragraph 1c of this Article shall be brought promptly before a judge or other magistrate authorised by law to exercise judicial functions and shall be entitled to trial within a reasonable time or to release during the procedure. Any person deprived of his liberty by arrest or detention shall have the right to appeal to a court to decide within a short time on the lawfulness of his detention and to order his release if the detention is unlawful.

⁷⁶¹ M. GIALUZ, *Quando l'intelligenza artificiale incontra il diritto penale*, 10.

⁷⁶² Which, in the case of the subject being arrested, is translated according to the English formula: “«shall be brought promptly before a judge»”.

⁷⁶³ On this point, M. GIALUZ, *Quando l'intelligenza artificiale incontra il diritto penale*, 10.

⁷⁶⁴ Thus on this subject, reference is made to the decisions of the Strasbourg Court, In the sense of the necessary presence of the defender, Corte EDU., 14th October 2010, *Brusco c. Francia*, § 45 and also Corte EDU, 28th October 1998, *Assenov e a. c. Bulgaria*, § 146; Corte EDU, 26th June 1991, *Letellier c. Francia*, § 35.

the guarantees of a fair trial⁷⁶⁵. Even if it is only mentioned in this part, it is considered that the canon of equality of the parties and of arms must be respected and, at the same time, the defence must be allowed full access to the investigative documents since, in the absence of knowledge of the data justifying the arrest, the right to bring an action for verification of legality (in the case of detention) would risk being reduced to a mere formality.

As a third declination of the guarantee of access to the court, reference is made to the duty to state reasons⁷⁶⁶, which will be discussed in more detail in the following paragraphs.

7.3 The risk of profiling: the use of big data and the invasive approach with individuals. Article 22 GDPR and its regulatory boundaries

On closer inspection, personal data can also be used for profiling individuals. In some cases, the aim and objective of data collection is just that. In other cases, however, profiling is 'obtained' or derived precisely from the misuse of such data. This risk is one of the key possible risks of using massive amounts of data.

The idea behind profiling is precisely that of 'extending information and data on individuals or groups of individuals in such a way that individual 'profiles' or propensities can be obtained.

The “Profiling⁷⁶⁷ is a technique of (partly) automated processing of personal and/or non-personal data, aimed at producing knowledge by inferring correlations from data in the form of profiles that can subsequently be applied as a basis for decision-making. [...]”⁷⁶⁸.

For example, if applied to the case of justice, the risk (as has been and is repeatedly emphasised among critics of the application of these instruments in the North American legal systems) also in the criminal sphere, when an assessment is to be made to analyse, elaborate or simply take into consideration certain characteristics of the subjects, is to create 'categories of individuals' and therefore this real 'profiling' can have positive or negative effects for the individual depending on the concrete cases.

⁷⁶⁵ Corte EDU, 31th January 2002, *Lanz c. Austria*, § 41; Corte EDU, 13th February 2001, *Schöps c. Germany*, § 44.

⁷⁶⁶ On this guarantee, established Strasbourg jurisprudence requires a statement of reasons for any ruling on the *status libertatis*: «justification for any period of detention, no matter how short, must be convincingly demonstrated by the authorities».

⁷⁶⁷ The modality through which profiling takes place is as follows: if certain individuals or groups of individuals present certain characteristics (X), the system will register that all individuals presenting this characteristic will fall into this 'group of individuals' either taken individually or considered 'as a group'. On the basis of the system these groups of individuals will be recognised as having a certain probability of possessing another characteristic, which for the sake of simplicity will be called Y.

⁷⁶⁸ F. BOSCO (et oths), *Profiling Technologies and Fundamental Rights and Values: Regulatory Challenges and Perspectives from European Data Protection Authorities*, in S. Gutwirth (et oths) (eds), *Reforming European Data Protection Law*, Berlin, 2015. See also M. HILDEBRANDT, *Profiling and AML*, in K. Rannenber (et oths) (eds), *The Future of Identity in the Information Society. Challenges and Opportunities*, Berlin, 2009.

In the case of the application of such instruments in certain procedural phases, or in the commensuration of punishment, the objective proposed therein would be to try to arrive at a choice on punitive treatment that responds mostly to needs that allow the judging body, as a support to the judge, to be able to evaluate more fully all the elements at its disposal in order to be able to take the decision with a 'more complete' assessment⁷⁶⁹.

On closer inspection, one can see how, the risk is that one moves from a kind of foresight to an influence that may lead to more or less advantageous choices for the individual himself, or even, in the worst cases, to forms of illegal and immoral manipulation.

Therefore, the debate and the issue is mostly centred on the need to find a way, even though it is not possible to completely eliminate the risk of profiling, since in any case a data collection takes place, to find a solution so that there is a selection of data that, although it entails a grouping, does not have as a derivative effect that of linking a greater propensity to crime or dangerousness, but can, for the most part, adhere to an assessment linked solely and exclusively to the individual.

As already mentioned, European Regulation 2016/679 (GDPR) lays down a specific provision on the processing of data and circumscribing the risk of profiling. In particular, Article 22 provides protection for the data subject who has the right not to be subjected to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or which affects him or her in a similar significant way. Indeed, it is precisely the fact that it is a decision taken not using other elements that is relevant. Indeed, if a decision having a significant impact on the data subject is reached 'solely' through the processing of data by an algorithm, then that provision applies including certain protections and guarantees afforded to the data subject.

On the other hand, if the algorithm merely aids and supports an individual's decision-making process, neither Article 22 nor its legal protection applies⁷⁷⁰.

In conclusion, clearly the risk of profiling has spin-offs that touch several spheres: from privacy, to the processing of personal data and until the risks of stereotyping, inequality, and discrimination due to the resulting 'classifications'⁷⁷¹.

⁷⁶⁹ G. SARTOR, *L'intelligenza artificiale e il diritto*, 69.

⁷⁷⁰As the former "Article 29 Data Protection Working Party" (now "European Data Protection Board" - EDPC/EDPB) explained in its Guidelines WP251[1], "an automated process produces what is in fact a recommendation concerning a data subject. If a human reviews the outcome of the automated process and takes other factors into account when making the final decision, that decision will not be based solely on the automated processing".

⁷⁷¹There are some states that have taken a stand and passed regulations on profiling techniques. In particular, The California Privacy Rights Act (CPRA)[2], passed on 4 November 2020, defines 'profiling' as 'any form of automated processing of personal information...'. The CPRA does not distinguish whether the actual decision was

7.3.1 *The collateral risks associated with profiling: the delicate drifts of stereotyping*

On closer inspection, in addition to the risks that have so far been mentioned arising from the risk of profiling, there are other aspects that are deemed necessary in order to be able to carry out a complete balancing act, which need to be assessed: the possible violations of privacy and data protection. At the same time, however, one must take into account how profiling may also trigger new risks of stereotyping, inequality and discrimination because of the classifications and categorisations on which they are based.

Indeed, as a negative effect of its degeneration, profiling risks leading to choices that undermine the interest of individuals to undergo fair and correct algorithmic treatment, i.e. the interest of not being subjected to unjustified prejudices as a result of automatic processing.

Indeed, the combination of big data and A.I. tools makes it possible to automate decision-making processes even in areas that require somewhat complex choices. They are such because they are choices based on numerous factors that are not exactly predetermined. No doubt, at first glance, however, these new types of decisions could bring benefits and improve the quality of certain decisions (not only in the legal sphere).

However, as already mentioned, they may carry several risks at the same time.

In other words, the explanation for this concern lies in the fact that the model that is built by the machine learning algorithm assigns individuals a score that reflects the probability that the individual has the characteristic predicted by the system (in the legal case, for example, that he or she has a higher degree of risk of committing a crime).

The resulting score allows a sort of classification to be made between individuals and creates a sort of label.

Finally, from the resulting label the algorithmic system will issue a certain decision that is closely linked to the classification produced by the system itself⁷⁷². Indeed, automatic systems are able to reveal propensities and make a sort of forward-looking assessment (which is why it is also called prognostic) that is sometimes more correct and more careful than that of a human. However, it must be considered that automatic systems can avoid human propensities to error, particularly in statistical inferences, as well as the prejudices -ethnic, social, gender, etc.- from

made by the algorithm or not. The mere fact that personal information has been processed by automated means, to produce a recommendation on how to act, provides additional protections and safeguards, whether the final decision is made by a human being or by a machine.

⁷⁷² In recent years, there has been a wide-ranging debate on the prospects and risks of algorithmic decisions. Some scholars have pointed out that in many areas, algorithmic predictions and decisions, including those concerning the evaluation of individuals, can be more accurate and effective than human ones. Thus on the point, G. SARTOR, *L'intelligenza artificiale e il diritto*, 69.

which we are often affected. More generally, automatic decisions can reduce the 'noise' inherent in human decisions (the fact that similar cases can have different decisions without reasonable explanation). It has been observed that in many areas -from investments to personnel recruitment, to the granting of probation- algorithmic determinations are better, with reference to the usual criteria, than those made even by experienced individuals. Others, however, have emphasised the potential for error and discrimination of algorithmic decisions. It is true that only in rare cases will algorithms make explicitly discriminatory decisions by implementing direct discrimination (disparate treatment), i.e. basing their predictions on prohibited characteristics such as race, ethnicity or gender. More often, the result of an algorithmic determination will result in indirect discrimination (disparate impact), i.e. it will have a disproportionately unfavourable impact on individuals belonging to certain groups, without an acceptable justification.

7.3.1.1 Discriminatory risk: can a machine's decisions be filtered by bias and discrimination?

Taking the starting question as a starting point, the question reflects the great fear and concern that some debates are igniting today. Especially in the United States, as anticipated in the previous paragraphs, one of the major pivots of the current debate on the use of such instruments has concerned precisely the possibility and consequently the risk that the discriminatory component they possess is ineradicable.

This is explained by the fact that systems that are based on learning methods (in particular those that are of interest here, i.e. supervised) learn from and make use of data processing. They therefore learn from the examples that are contained in the 'training period'. They work with existing data and elements that may undoubtedly have merits and demerits of the exemplified behaviour and thus convey not only inaccuracies but also errors and biases.

This risk could occur whenever there is a direct correlation between discriminatory characteristics and certain input data that are used by the system⁷⁷³.

However, injustice and 'unfairness' in the data could also result from the use of data based on biases, biased human judgements and evaluations.

Finally, other errors, and these are the ones that could perhaps most hardly be predicted, derive from data that should reflect characteristics in population statistics but in reality, do not.

⁷⁷³ "Suppose, for example, that a human resources manager of a company never hired candidates of a certain ethnicity because of his bias and that individuals belonging to that ethnicity mostly live in certain neighbourhoods of the city. A set of training based on that manager's decisions would teach the system not to select individuals residing in those neighbourhoods, resulting in a failure to hire," see G. SARTOR, *L'intelligenza artificiale e il diritto*, 72 s.

For example, this risk would be possible when certain individuals who are part of certain groups subject to certain controls are more likely to be regarded as perpetrators of criminal behaviour (whether they have actually engaged in it or not)⁷⁷⁴.

The first risks leading to deterministic and discriminatory effects have already been partly illustrated when describing the decisions in the US context that led to the first critical reflections on the subject⁷⁷⁵.

Now, as already noted in the previous chapters, a concrete risk that is closely connected to the use of algorithms is the projection of racial prejudice in the result provided.

Therefore, a distinction is made between 'derived' and 'autonomous' biases. This distinction depends on whether the bias is located in the source code, which could then reflect the biases either peculiar to the programmer or generated by the reference values of the organisation in which the programmer operates (e.g. if one thinks of the inclusion or exclusion of characters that identify a certain category of subjects); one can imagine that such effects are not so much related to the source code and the person who programmed the algorithm, but rather to the data that are fed into it to elaborate the statistical models, since the algorithm uses training data to feed itself. On closer inspection, such data could therefore reflect bias on the part of the person who selected it or contain errors due to generalisations based on incomplete data, inadequate methods of collection⁷⁷⁶, biased data, inconsistent data, or data that are not adequately representative of the minorities that are involved⁷⁷⁷. In other words, if the sample of data collected is thus significantly larger for a certain group of individuals (e.g. African-Americans) and much smaller for another, then the over-represented (or under-represented) group is consequently disadvantaged by the misrepresentation of reality. Lastly, another possibility relates to the risk that in the absence of specific information or corrections during programming, the algorithm might autonomously identify certain characteristics that refer to protected categories, associating their holders with detrimental treatment (this would be the most problematic case because it would be unsupported and out of human control).

⁷⁷⁴ Clearly, the most striking case here is one where this example is most evident because of a problem related to inequality and disproportionality; this is an issue that has been the subject of debate in the United States for several years. Suppose, for instance, that in the evaluation of applications for parole, the presence of a criminal record against the offender weighs unfavourably.

⁷⁷⁵ It is no coincidence, in fact, that in the second appendix to the Ethics Charter, precisely in view of the discriminatory and deterministic effects that risk assessment tools have had in the United States, these devices are placed in the category of those whose use is desired with the most extreme reservations.

⁷⁷⁶ Consider the case where police files are used as a way of acquiring data where the percentage of immigrants or African-Americans on file is particularly high: the system would learn that immigrants or African-Americans are more likely to commit crimes.

⁷⁷⁷ QUINTARELLI S., *Intelligenza Artificiale: cos'è davvero, come funziona, che effetti avrà*, Turin, 2020, 96.

On closer inspection, the data for learning the algorithm are manually labelled by humans; it is evident that, even if unconsciously, there is a risk that value judgements and social preconceptions may be reflected in this activity.

In other words, and this would be the most important and probably also the most difficult step to take, detecting and correcting such distortions would not be entirely possible since we would return to the problem of algorithmic transparency and the controllability of the algorithm's actions. Moreover, it would be quite difficult to establish *ex ante* whether or not a given algorithm performs and realises discrimination, since this will emerge only when individuals complain that they were victims of discrimination.

Therefore, it is precisely the impossibility of a full *ex ante* control due to the opacity flaw of the algorithm that would not allow a full control over its functioning. To translate the discourse to the level of justice, risk assessment systems are based on historical criminal justice data. The quality of these projections depends on the quality of the criminal justice system data that were used to develop them. Consequently, persistent problems with the effects of mass incarceration, discriminatory policing practices and other discrimination in the criminal justice system ultimately lead risk assessments to project the same discrimination and prejudice into the future⁷⁷⁸.

In conclusion, from the reflection carried out, it can be seen that the much sought-after neutrality of algorithms probably represents an unattainable step in itself since it is almost impossible to imagine that the creators of algorithms would not be able to influence the systems with their own values. Lastly, it is noted that even if one were able to identify the source of the

⁷⁷⁸ This problematic and risky situation has already been reiterated and analysed in May 2016 ProPublica published a study entitled 'Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks' showing how the results provided by the algorithm were systematically racially biased, resulting in higher rates of danger in African-Americans than whites, with African-Americans more than twice as likely to be identified as "high risk" and 45% more likely to commit any type of crime in the future, rising to 77.3% when assessing the risk of violent recidivism. White defendants, on the other hand, were more likely (approximately 63%) to be labelled as 'low risk', but then reoffended in the following two years. The ProPublica team of journalists compared the risk scores assigned by COMPAS with the defendants' actual reoffending in the following two years (finding an accuracy level of 61% in the case of general reoffending and 21% in the case of violent reoffending). By dividing the population into black and white individuals and comparing the two samples, the team found that for black individuals, the algorithm predicted a high number of what are statistically termed false positives, i.e. individuals classified as high risk who did not commit a new crime in the next two years. The company Northpointe, Inc. responded by commenting that 'Northpointe does not agree that the results of your analysis, or the claims being made based upon that analysis, are correct or that they accurately reflect the outcomes from the application of the model' and by specifying that the algorithm produced the predictions with the same accuracy for both samples analysed (predictive parity). As ProPublica noted, the algorithm should, in any case, identify high-risk individuals regardless of their ethnicity. The bias found by ProPublica's analysis proved to be related to the use of judicial precedents (mostly unfavourable for black convicts), which had led the system to overestimate the risk of recidivism for African-Americans. See on this point A. VESPIGNANI, *L'Algoritmo e l'oracolo: come la scienza predice il futuro e ci aiuta a cambiarlo*, Milan, 2019, 106 ss.

bias, precisely because of the algorithm's complex learning mechanism, it would be impossible to succeed in 're-educating' it correctly. In fact, behind the apparent impersonal or objective 'efficient façade', algorithmic systems instead reflect the intentions of those who design or commission them, thus generating an operational and asymmetrical power over other people⁷⁷⁹.

7.3.1.1.1 *What solution? Data cleansing and constant monitoring*

In addition to identifying the possible problems and risks associated with such introduction, there is undoubtedly a need to overcome these issues.

One of the situations that could be proposed would certainly involve work that calls several actors into play. Indeed, one would first have to select, and here we refer to the following pages, the choice of data for the algorithm process.

In particular, it could already be a starting point, in order to reduce risks and discriminatory drifts, to work on the not only accurate but also *ex ante* choice of the data to be taken into consideration. In fact, it would be desirable and feasible to exclude information such as gender, race and other data that are historically linked to social segregation dynamics, or to devise a sort of 'filtering system', carefully selecting the data to be entered into the system and favouring only the so-called 'neutral' data that cannot categorise the subject.

This step alone, however, would not be sufficient since, once the data has been identified, the algorithm, in order to create its learning model and to be able to issue its evaluations through the output score, would have to compare itself with the data that is already inside it and is compared and analysed by it. Therefore, it should be possible to ensure that the previous data is 'cleaned' of judicial errors. In doing so, however, one would have to entertain the notion that 'cleaning up' the comparison data altogether would be a rather difficult objective to achieve.

However, while it is true that algorithmic systems, especially those based on machine learning, may reproduce or even exacerbate existing inequities, it is also true that algorithmic processes may be more controllable than a human decision, whose motivation, if not well worked out, is also hardly controllable. Moreover, as a further favourable element, they have the possibility of 'improving over time'.

⁷⁷⁹ E. SADIN, *Critica della ragione artificiale: una difesa dell'umanità*, Rome, 2019.

This observation is shared by criminologist Aleš Završnik, who points out that the construction and interpretation phases of algorithms are 'produced by men for men and, however they are conceived, they cannot escape human errors, prejudices, human interests and human representation of the world'.

8 *The problem of discrimination overcome by the rationality of the machine?*

Indeed, as the first questions and debates that arose following the Loomis case in the US have already shown, one can see how the first applications of these instruments are useful because they can act as true litmus tests that highlight the most relevant doubts and questions that have arisen in the margins of the first applications.

Indeed, the numerous questions left unresolved by the Loomis case show how, even the position of the Wisconsin Court seems to be not entirely decided: it would almost appear that it is content with the circumstance that the judges on the merits, at least formally, did not take into account the discriminatory effects resulting from an individual's belonging to a group (in particular that of male sex offenders).

However, by doing so, one ends up underestimating the real extent of the problem; indeed, although the risk score is determined by processing the data collected *hic et nunc*, the fact remains that the subject is framed in a socio-criminal profile that is based on the rate of recidivism in similar cases.

Indeed, the main objection that would seem to move from the first applications of such tools concerns the inclusion, among the various variables relevant to the determination of the level of risk, of demographic, family socio-economic factors that contribute to characterising as more dangerous individuals those belonging to certain minorities or social classes⁷⁸⁰.

Indeed, predictive algorithms use big data on criminal records over the past decades in order to categorise criminals into different groups and subgroups to which a risk value is attributed. The data provided as input serve to describe the profile of the subject and then combine to determine the summation of the different scores (positive or negative) referred to the individual categories; however, the output is contaminated by the historical trend towards deterrent treatment and prejudice against certain criminal figures.

The main question concerns the fact that, in order to be able to try to rationalise the discriminatory risky effects of data that are tainted *ab origine*, one should try to ensure greater reliability; this is extremely difficult to achieve, since the software draws on a very extensive database that also includes historical periods in which, from common experience, there was a tendency towards ethnic discrimination and the subjectivisation of punishment.

⁷⁸⁰ COMPAS, for example, takes into account criminal convictions of the defendant's parents, the use of toxic substances or drugs, or any crimes the household members have been victims of in the past. LSI-R also considers among the relevant factors the involvement of social services, grades attained in high school, the chances of finding a job with a good salary, the crime rate in the neighbourhood where the person lives or grew up. See S. B., STARR, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 813.

The statistical investigations that have been conducted in recent years have shown how precisely the variables of a socio-economic nature, relating to ethnic origin or the degree of schooling⁷⁸¹ are often a determining factor in measuring the risk of reoffending. However, another question undoubtedly connected and deriving from the first one relates to the fact that precisely because the weight that the predictive algorithms attribute to the factors de quibus in the overall assessment is not known, it could be assumed that a given subject belonging to a 'risk' category is considered more dangerous on the basis of mere generalisations (the so-called *group-based generalisations*).

On closer inspection, since the discriminatory effect produced by the socio-economic variables is thus quite evident, some authors suggest precisely to expunge them from the risk assessment parameters, trying to mimic the analysis only to the criminal record, the age of the offender's first arrest and the characteristics of the crime committed⁷⁸².

8.1 *The problem of data transparency and the opacity of A.I. systems*

Undoubtably, this is the first *cornerstone* of the theoretical landscape in this research. In fact, “the need for accessibility is the quintessential feature in the discourse about the use of algorithms in decision-making processes, both in a private and public context”⁷⁸³.

Indeed, as already reiterated in the course of the paper, in relation to the level of problematicness arising from the (probably inevitable) *opacity* that accompanies digital risk assessment tools, here, unable to find a solution to this problem, an attempt is made to formulate an optimistic approach that looks to the future and calls for action not only by individual legal systems but also by the European context.

The greatest difficulty in the face of the 'impenetrability' of the machine is exacerbated by the use of programming language: even if the source code were known, the judge could not know how the assumptions of human reasoning were translated into code by the developers, nor could he or she verify the correctness of these assumptions by himself or herself.

What must be considered is that it is precisely from the jurisprudence of the European Court of Human Rights that a kind of evidentiary exclusionary rule descends, which could close the door to all evidence that is automatically generated and that is not susceptible to *ex post*

⁷⁸¹ According to a recent study, individuals who have not completed higher education are 47 times more dangerous than those who have graduated.

⁷⁸² S. B., STARR, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 850; D. KEHL – P. GUO – S. KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing, Responsive Communities Initiative*, 25

⁷⁸³ *Ibidem*.

verification. This reading is given and found from the reading of the same jurisprudence that anchors the fairness of the trial, as a whole, to the equality of arms⁷⁸⁴.

The necessary requirement of transparency becomes even more relevant when it concerns a stage of the judicial decision-making process and, in particular, the criminal trial.

On closer inspection, it should first be noted that the concept of transparency is a concept that does not have its own unambiguous definition and must necessarily be compared and discussed with the concepts of '*explanability*' and '*justifiability*'⁷⁸⁵. Following the concept in this sense, the actual transparency in the first place undoubtedly depends on the precision of the scientific theory and secondly on the clarity of the language used to translate it into a mathematical formula⁷⁸⁶.

Lastly, therefore, it is arguably necessary to ensure the transparency of the evaluation process carried out by the instrument itself and the consequent possibility of challenging the reliability of the output. To this end, it would perhaps be preferable to directly involve public agencies and the scientific community⁷⁸⁷ in order to build transparent systems that respect the canons set by the European Ethical Charter for the use of artificial intelligence⁷⁸⁸.

Indeed, the problem with the result obtained from a risk assessment or a predictive algorithm, as already mentioned, clearly hangs on the data with which the algorithm is trained⁷⁸⁹. Therefore, it is necessary, as mentioned earlier, to understand where such data is

⁷⁸⁴ On this point, reference is made to ECHR, 28.8.1991, *Brandstetter v. Austria*, § 68. However, for a complete overview; S. QUATTROCOLO, *Artificial intelligence*, 148 ss.

⁷⁸⁵ Justice Abrahamson's concurring opinion clearly highlights the lack of transparency in the functioning of predictive algorithms. The greatest concern is the inability of judges to know precisely what input factors are considered by the software, and how these factors are weighted against each other. The awkwardness in the face of the machine's 'impenetrability' is exacerbated by the use of programming language: even if the source code were known, the judge could not know how the assumptions of human reasoning were translated into code by the developers, nor could he or she verify the correctness of those assumptions himself or herself. There also remains the problem of the compatibility of the use of tools designed for pre-trial decisions to sentencing. When a tool is developed for one context, such as risk assessment for the granting of alternative measures to detention, it may not necessarily be automatically adapted for sentencing. The obscure aura that surrounds the revelatory indexes of dangerousness (and the weight attributed to them) significantly limits the right of defence of the defendant who is not put in a position to scrutinise the output of the assessment and to verify its (even only) formal correctness. On closer inspection, in a procedural system governed by the principle of adversarial evidence-gathering, full discovery should be ensured on all evidence used against the defendant. It is therefore astonishing that Judge Abrahamson's arguments only found a place in a separate opinion that did not disagree with the decision taken by the panel. Having said this on the US experience, let us now turn our gaze to the Italian system to examine whether risk assessment tools and techniques for the commensuration of punishment can be used within the criminal trial.

⁷⁸⁶ Così sul punto C. CHESSMAN, *A 'source' of Error: Computer Code, Criminal Defendants, and the Constitution*, California, 2017, 2016.

⁷⁸⁷ See, M. GIALUZ, *Quando l'intelligenza artificiale incontra il diritto penale*, 21.

⁷⁸⁸ Cfr. anche i suggerimenti prospettati, soprattutto per evitare le discriminazioni basate sulla razza e sulla condizione sociale, C. DOYLE – C. BAINS – B. HOPKINS, *Principles of pretrial release: reforming bail without repeating its harms*, in *The Journal of Criminal Law and Criminology*, 1973, Vol 108, No. 4, 17.

⁷⁸⁹ The indications contained in Gruppo di Esperti MISE sull'intelligenza artificiale, *Proposte per una strategia italiana per l'intelligenza artificiale*, 2 July 2020, 11.

obtained and extracted from, and then it must be collected in a transparent and legitimate manner⁷⁹⁰.

The second point that is critically emphasised is who gets to decide and intervene in the choice of data. This is critical because it affects reliability, since the data, if already flawed, will invalidate the result by transmitting all the bias.

Indeed, the principle of algorithmic transparency can be qualified as 'the obligation, incumbent on those who make decisions with the aid of automated data processing systems, to provide the recipients with an understandable explanation of the procedures used and to justify the decisions taken in this respect. In other words, to fully guarantee the decoding of such tools, the information must not only be accessible, but also comprehensible⁷⁹¹.

In particular, Article 15 of the GDPR itself marks within the principle of transparency the need for language that is simple and comprehensible to be used. The principle of 'comprehensibility' is thus connected to the right of access to the algorithm, i.e. the guarantee of the widest possible access to algorithmic information, thus being able to identify both 'its authors, the procedure used for its elaboration, and the decision mechanism, including the priorities assigned in the evaluation and decision-making process and the data selected as relevant'.

As already mentioned, technical opacity indicates the situation where the result of the algorithm is either not knowable or not comprehensible. With reference to models whose external behaviour (output) can only be described, but whose internal functioning cannot be known by reconstructing the logical path that led to the result, the expression 'black boxes' has been coined⁷⁹².

⁷⁹⁰ The issue was raised loudly after the Clearview case in the United States and also concerned the HART risk assessment applied in England a few years earlier.

⁷⁹¹ As noted on the point «transparency is not enough, in itself: transparency must be meaningful; the disclosure of the source code is not considered true transparency, because only experts can understand it». S. QUATTROCOLO - C. ANGLANO - M. CANONICO - M. GUAZZONE, in *Technical Solutions for Legal Challenges: Equality of Arms in Criminal Proceedings in Global Jurist*, 2020.

⁷⁹² S. QUATTROCOLO, *Processo penale e rivoluzione digitale: da ossimoro a endiadi?*, in *Medialaws*, 3/2020. Furthermore, in this case, «the input and ultimate output of the system are observable, but how the system arrives at that outcome is unknown, even to those who created it. On this point, also, L. TILLER, *A Minority Report: The Unregulated Business of Automating the Criminal Justice System* in *The Business, Entrepreneurship & Tax Law Review's B.E.T.R. White Paper*, March 2019, 10 ss.

On closer inspection, the increasing reliance on big data to make any kind of decision has exacerbated the problem even further⁷⁹³. One can thus recognise three main circumstances that can lead to this flaw⁷⁹⁴:

- Opacity as an objective 'intentionally' pursued by the trade secret policies adopted by private companies that thus make the technical specifications of the algorithm and the source code inaccessible. This is in order to maintain competitive advantages over competitors on the market⁷⁹⁵;

- Opacity due to the technical expertise required for the intelligibility of the result, which is beyond the reach of ordinary citizens. Furthermore, if the algorithm design is not transparent, it may also be impossible to verify the reliability of the output for anyone other than the designer of the source code itself⁷⁹⁶.

- Intrinsic opacity in machine learning systems: such systems, in fact, operate according to a so-called 'deductive' logic, since they evolve and learn from data that are input from time to time. The direct consequence is that even when the source code is revealed, the reasons and steps followed by the machine to provide that particular result may not be fully comprehensible (so-called intrinsic opacity).

8.1.1 *Current scenario and possible future solutions*

On closer inspection, the problem has already been advanced in the US judicial system where the use of algorithms in the *sentencing phase* is still widespread, since it is believed that the possibility of challenging the result (by having recourse to the algorithm's user manual) would in itself be sufficient to guarantee the right of defence⁷⁹⁷.

However, just such an argument thus taken out of context cannot entirely convince, especially if one also takes into account the 'psychological weight' that the result of an algorithm that does not account for the data but reconstructs them by mysterious means into a 'pre-packaged decision-making package with a given solution' can exert on the adjudicating body⁷⁹⁸.

⁷⁹³ So much so that to describe the pervasiveness of the phenomenon, one speaks of a 'black box society'. So much so that to describe the pervasiveness of the phenomenon one speaks of a 'black box society'.

⁷⁹⁴ J. BURRELL, *How the machine 'thinks': Understanding opacity in machine learning algorithms*, in *Big Data & Society*, June 2016, 1 e ss.

⁷⁹⁵ Very often, this is referred to as 'proprietary protection' or 'corporate secrecy'.

⁷⁹⁶ S. QUATTROCOLO, *Equità del processo penale e automated evidence alla luce della Convenzione europea dei diritti dell'uomo*, in *Revista Ítalo-Española de Derecho Procesal*, Vol. 1, 2019.

⁷⁹⁷ C. CESARI, *Editoriale: L'impatto delle nuove tecnologie sulla giustizia penale – un orizzonte denso di incognite*, in *Revista brasileira de direito processual penal*, Porto Alegre, Vol. 4, No. 3, 1177 ss.

⁷⁹⁸ *Ibidem*.

In other words, the right of defence would in itself imply a series of guarantees, including the possibility of verification of any useful contribution to influence the judicial decision. Following this, the opacity flaw caused by the secrecy of the software precludes the verification of the results that should in any case always be ensured.

It appears evident that such delicate phases of the trial cannot in any case be regulated by a sort of '*black box*'⁷⁹⁹. The risk, therefore, which would be incurred, as analysed above, would also consist in the dramatic effect that the subject would not be put in a position to fully exercise his right of defence and, at the same time, would not be able to notice any discriminatory flaws produced by the software itself.

The solution, therefore, could be sought in the effort to elaborate clear, precise rules and a '*hard*' core of information that must be extruded to ensure the fairness of the process.

Having thus illustrated the different forms and species of opacity, the solution, in the first case, would be to make the source code available, making it '*open source code*' of the algorithm in order to be able to contest and consequently correct any manipulations, distortions or errors⁸⁰⁰. However, even this solution would seem to have weaknesses and would struggle to become a real remedy, since then only experts in the field would be able to understand the meaning of the code.

Indeed, full comprehensibility would be ensured as long as it is accompanied by explanations that translate it into the underlying '*legal rule*' and make it readable and understandable, both for citizens and for the judge.

Another solution could consist in the use of an independent expert who could assess and verify the reliability of the algorithmic result *ex post*.

Lastly, to obviate the same issue, the hypothesis of developing so-called '*zero-knowledge proof*' encryption software has also been put forward, with which it would be possible to identify the criteria that govern the algorithm's policy, but without having to reveal the policy itself in order to verify the correctness of the output.

8.1.2 *The risk of undermining the guarantee of the 'equality of arms at trial': the right of access for the defendant*

On closer inspection, it seems first of all relevant to take up the position expressed by the European Court of Human Rights, according to which '*equality of arms*' means '*equivalence of*

⁷⁹⁹ Because 'if it were not even possible to know this content for intellectual property reasons, then the right of defence would cease to exist', J.NIEVA-FENOLL, *Intelligenza artificiale e processo*, Turin, 2018.

⁸⁰⁰ This solution is also suggested by the Ethics Charter, which suggests, for example, as a remedy the creation of authorities in charge of the verification and certification of the automatic models used in the process.

the chances' offered to each party to the proceedings to be able to convince the court of its view of the facts at issue in the proceedings⁸⁰¹.

Indeed, it is precisely by virtue of the general canon of due process and, more specifically, the principle of equality of arms, that even the admission and assessment of automatically generated evidence appears to be potentially at odds with fundamental guarantees of the convention⁸⁰².

Indeed, in the analysis we are interested in, it would be necessary to assess whether one of the cornerstones of due process would be undermined if such instruments were introduced.

On closer inspection, the reasoning should be subdivided along two lines: firstly, it should be noted primarily that if it is decided to use and introduce risk assessment in respect of the defendant, but without it being otherwise possible to verify *ex post* how such results were achieved, the defence would be deprived of arguments to challenge the reliability of the risk calculation; this would therefore run counter to the principle of equality of arms.

Consequently, such an impossibility of *ex post* verification could derive as much from an inherent feature of the software as from the existence of a trade secret.

In both cases, however, the solution would inevitably seem to be to have to forego the evidence which, if generated by a black box, would in any case make an effective criticism of its reliability impossible⁸⁰³.

On closer inspection, the fundamental prerequisite for a fair trial is that the defendant be put in a position to adequately prepare his or her defence strategy: in particular, the defendant is only able to protect himself or herself fully and effectively if he or she is put in a position to have access to all the appropriate elements that form the basis of the judicial decision. In fact, the essential paradigm of equality of arms is represented by the possibility of presenting one's arguments in conditions that do not disadvantage one party with respect to the others. Precisely

⁸⁰¹ In this sense, in particular, ECHR, 12.4.2006, *Martinie v. France*, § 46, although not referring to criminal proceedings.

⁸⁰² It is well known that first and foremost and even in the absence of an explicit enunciation in the text of Article 6 ECHR, the principle of equality of arms has been modelled by the Court's jurisprudence as the lintel, together with the related canon of cross-examination, of procedural fairness as a whole. M. CHIAVARIO, *Art. 6*, in S. Bartole - B. Conforti - G. Raimondi (eds), *Commentario alla Convenzione Europea dei diritti umani e libertà fondamentali*, Padua, 2002, 192. B. CONFORTI - G. RAIMONDI, *Commentario alla convenzione europea dei diritti dell'uomo e delle libertà fondamentali*, Padua, 2002, 192. This is a long-standing acquisition in Strasbourg jurisprudence, on which see already ECHR, *Neumeister v. Austria*, Rec. 1936/63 (1968), § 22 of the reasons in law, which recognises equality of arms as a feature of the fair trial, on the basis of numerous earlier decisions and opinions of the European Commission, which was then in charge of conducting an access filter at the Court; CEDU, *Delcourt v. Belgium*, ric. 2689/65, 1970, § 28: «The principle of equality of arms does not exhaust the contents of this paragraph; it is only one feature of the wider concept of fair trial by an independent and impartial tribunal».

⁸⁰³ S. QUATTROCOLO, *Sui rapporti tra pena, prevenzione del reato e prova nell'era dei modelli computazionali psico-criminologici*, in *Teoria e critica della regolazione sociale*.

for this reason, from a procedural point of view, but with implications and repercussions also on the substantive level, the right of "access to the algorithm", understood as the right to access the logic and specific functioning of the same, becomes relevant: the result provided by the algorithm is assimilable to an expert opinion and, as such, can be included among the means of procedural evidence, capable of providing elements to be used as a basis for the decision⁸⁰⁴.

Indeed, the greatest problem arises when the possible impossibility of access to the source code or of actually being able to understand the operation of the black box that generated a given result that is considered by the adjudicating body, determines an implicit risk to the equality of arms. If the essence of procedural fairness resides in the full right to be able to try to convince the judge of one's reconstruction of the facts by effective means, even by challenging the admissibility and accuracy of the evidence, the impossibility of verifying the output of an algorithm a posteriori may represent in nuce a violation of Article 6(1) ECHR (regardless of the existence of an upstream violation of the right to privacy).

Because of the opacity defect, the software of the algorithm is not scrutinisable: this, consequently, entails the use in Court of a result whose generative process is neither known nor accessible and therefore can be classified as the disclosure of evidentiary material necessary for the preparation of the defence⁸⁰⁵.

On this issue, therefore, it must be further considered that the guarantees of access to the evidence have now been recognised by the European Court⁸⁰⁶ for any type of evidence, including computers and files relevant to the charges against the accused⁸⁰⁷.

⁸⁰⁴ The principle is violated where "he is denied the opportunity to attend the proceedings, or where he is unable properly to instruct his legal representative". In *Kuopila v. Finlanda*, the Court held that the failure to disclose evidence to the defence constituted a violation of the principle of equality of arms. The defence was precluded from cross-examining a supplementary police report. On this point, the Court observed that «the procedure did not enable the applicant to participate properly and in conformity with the principle of equality of arms in the proceedings before the Court of Appeal».

⁸⁰⁵ In particular, according to the Court, "the failure to disclose material evidence, including elements that could have led to an acquittal or a reduction of the sentence, may constitute a denial of the facilities necessary for the preparation of the defence, and thus a violation of Article 6 § 3(b) of the Convention". See, Eur. Court of human rights, 4th Section, 31.3.2009, 21022/04, *Natunen v. Finland*, § 43 «. Failure to disclose to the defence material evidence, which contains such particulars which could enable the accused to exonerate himself or have his sentence reduced would constitute a refusal of facilities necessary for the preparation of the defence, and therefore a violation of the right guaranteed in Article 6 § 3 (b) of the Convention».

⁸⁰⁶ Eur. Court of human rights 1st Section, 9.5.2003, 59506/00, *Georgios Papageorgiou v. Greece*, §37.

⁸⁰⁷ For example, in *Georgios Papageorgiou v Greece*, the Court found a violation of the applicant's right to a fair trial precisely because the request for the production of extracts of computer files (functional to the defence), deemed unnecessary by the Athens Court of Appeal, had been refused. Thus, the right of access to the algorithm can be brought within the scope of the ECHR's guarantees of the right to have the necessary facilities to prepare one's defence. As noted in the case *Gregičević v. Croatia*, «The accused must have the opportunity to organise his defence in an appropriate way and without restriction as to the ability to put all relevant defence arguments before the trial court and thus to influence the outcome of the proceedings». On this point also *Mayzit v. Russia*, no. 63378/00, § 78, 20 January 2005; *Connolly v. the United Kingdom* (dec.), no. 27245/95, 26 June 1996; *Can v. Austria*, no. 9300/81

The accused, therefore, must be able to obtain copies of all documents relevant to the trial; it is therefore also possible to bring within the scope of this guarantee the possibility of access to the source code and specifications of the algorithmic software⁸⁰⁸.

Therefore, an essential condition for the full respect of this principle is the accessibility, i.e. the knowledge of the evidence and arguments put forward by the opposing parties in court. Indeed, it is precisely the opacity of algorithms that prevents such dialectics, surreptitiously admitting the entry into court of a contribution that cannot be disputed by the defence. Thus, the very introduction of algorithmic evidence into the trial entails the potential violation of fair trial.

Apart from these considerations, therefore, it is evident how the opacity of the algorithmic decision-making process undermines *in nuce* the right of defence, which implies being able to effectively conduct all arguments and evidence in favour⁸⁰⁹. The parties would thus be precluded from verifying the accuracy of the data and contesting the algorithmic result effectively, as they would not be able to know the steps that led to a given result, the quality of the data entered and the weight given to them, the source code and its technical specificities⁸¹⁰. In other words, the evidence (or more generally, the elements that may influence the final commensuration of the sentence) generated by software or computational systems, prevent the defence from validating its genesis and thus the genuineness of the data⁸¹¹.

In fact, the argument based on the output of an unscrutinisable algorithm becomes an "argument that can only be used by one party, since the difficulty of explaining its genesis becomes a difficulty of contesting its reliability.

The impossibility for the defence to challenge the accuracy and therefore also the reliability of the evidence against the accused produces a strong imbalance of power between prosecution and defence. Indeed, the potential presence of a discriminatory bias in the algorithm consequently causes harm to the minority that is the object of the bias, most of the time unaware of suffering injustice due to the black box effect.

Clearly, although the analysis carried out would seem to stop at an embryonic stage, nevertheless, there are possible perspectives for overcoming the problems outlined, which

⁸⁰⁸ See, *Rasmussen v. Poland*, §§ 48-49; *Moiseyev v. Russia*, §§ 213-218; *Matyjek v. Poland*, § 59; *Seleznev v. Russia*, §§ 64-69.

⁸⁰⁹ J. NIEVA- FENOLL, *Intelligenza artificiale e processo*, 129

⁸¹⁰ As noted in the Loomis case, the Wisconsin court denied the problem of accessibility to the algorithm, arguing that the COMPAS result was not the only element on which the court based its decision: the answer is inadequate because it allows the insertion of an element of evidence that cannot be scrutinised by the parties.

⁸¹¹ Così, S. QUATTROCOLO, *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale "predittiva"*, 1761 ss.

could therefore contribute to bringing the respect for the principle of 'equality of arms' back on an even keel.

8.1.3 A possible solution to overcome the obstacle of obscurity: towards greater transparency

Another problem that emerges in the use of algorithms is given by the lack of transparency of the method used by the predictive algorithm, - of the way in which the information already present in the calculation and that relating to the individual case are connected -, which is ordinarily a proprietary algorithm, that is, covered by copyright and, therefore, not knowable by the Courts, nor controllable by the defence, with the consequence that there is a lack of adversarial debate on the admissibility of the use of the tool and its results, with violation of the principle of due process, the presumption of innocence and the rights of the defence⁸¹².

As already mentioned, one of the major problems that at the same time also represents the possible solution to the problem of the opacity of algorithmic and computational processes is precisely transparency. However, in the sphere of automated data processing, transparency seems to have become the sole and determining parameter of the legitimacy of processing, deviously replacing the canon of legality. If software is designed according to parameters of transparency, the possibility of validation or falsification of its outputs is higher, and the GDPR, which recently came into force, and to some extent also the EU Directive 2016/680, on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection and prosecution of criminal offences or the execution of criminal penalties (Art. 20), seem to be inspired by this assumption⁸¹³.

However, this would still be 'mediated' transparency by the expert⁸¹⁴. In other words, it can be achieved by obtaining access to the source code, inputs and outputs of the software. In spite of this, however, such access does not guarantee a general understanding of the process that generated the output, because only computer experts may be able, and not always (see below), to derive meaningful and comprehensible elements from it

On closer inspection, one possible solution that could be proposed would be to create independent public authorities that would assess and certify the tools a priori and constantly monitor their functioning.

⁸¹² See A. M. MAUGERI, *L'uso di algoritmi predittivi per accertare la pericolosità sociale*, 19 ss.

⁸¹³ Recently transposed also in Italy with Legislative Decree 51/20184.

⁸¹⁴ A. KOENE - H. WEBB - M. PATEL, *First UnBias Stakeholders workshop*, 2017, in unbias.wp.horizon.ac.uk.

In fact, since it would not be possible within the justice system to ensure that a committee of experts or a special office would be in charge of scrutinising the functioning and processing of data in accordance with the data processing legislation and the soft law instruments indicated at European level, one could try to imagine the introduction of extra-processing bodies that could take care, at public level, of authorising, controlling and then certifying these instruments.

As has already been pointed out, access to the specific characteristics of the algorithm and the source code is at present unlikely due to policies that still make software private and therefore covered by industrial secrecy.

Surely a possible solution to overcome this issue could be to make accessible a certain amount of predetermined information, for example, which variables are used, for which objective the algorithm has been optimised, the type and quantity of data entered, the way in which the algorithm's performance is monitored, how the algorithm itself evolves over time, the factors relevant to the functioning of the algorithm itself, the data entered for its 'training', their classification and the weight attributed to each of them⁸¹⁵.

It follows that for the algorithm to be validly used in criminal proceedings, it is necessary to guarantee the very transparency that makes it auditable: "In this sense, effective transparency depends, firstly, on the precision of the underlying scientific theory and, secondly, on the clarity of the language used to translate it into a mathematical formula [...] a clear mathematical language allows an *ex post* reviewer to understand how the process has evolved from input to output"⁸¹⁶. In order to avoid overestimation effects and/or risks of 'false positives', and to allow the defence to verify the scientificity and accuracy of an 'enigmatic database' or data generated by a given computational process, an expert evaluation of the algorithm should be allowed - not unlike any scientific acquisition that enters the criminal trial - and in any case its empirical soundness should be assessed in an adversarial manner, respecting the rights of the defence⁸¹⁷.

Indeed, some scholars consider transparency the antidote to the discriminatory character of the algorithm. They assume that it is not possible to guarantee the neutrality of the algorithm in view of the intrinsic risk inherent in algorithmic decision-making, since they consider that the need to use data that can only ever be partial leads to a situation of inevitable bias, in the

⁸¹⁵ L. MCGREGOR - D. MURRAY – V. NG., *International human rights law as a framework for algorithmic accountability in International and Comparative Law Quarterly*, Vol. 68, 2019, 309-343.

⁸¹⁶ However, to fulfil the need for justification, the underlying scientific theory must be valid enough to provide a causal relationship between the input data set and the results. In the context of this study, we refer to the concept of 'validity' in the sense provided by the US Supreme Court in the aforementioned Daubert case"; thus S. QUATTROCOLO, *Artificial intelligence*, 17 s.

⁸¹⁷ V. MANES, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, 16.

sense that it would be impossible to provide a neutral representation of the initial information; it is not considered possible "to demand neutrality from a system designed to choose, filter or order information according to certain principles and conceived precisely in order to reduce uncertainty in a universe where the abundance of data does not allow for choice"⁸¹⁸.

Therefore, it is preferable, in the face of a vain neutrality, to demand the loyalty of the algorithm and that is to say, its transparency, making the criteria that preside over the algorithmic decision and the functioning of the system knowable, allowing access to the reasoning phases followed by the system in order to take a given decision; all this in order to avoid that the partiality of the knowledge on which the decision is based may conceal an unidentifiable discrimination. This is the characteristic of explicability⁸¹⁹, which must characterise A.I. systems that meet the requirement of transparency. The A.I. system will meet the criterion of explicability if the producer is able to anticipate the effects to which use of the system may lead, so that the user can identify them in good time and, if necessary, report any anomalies. Naturally, the level of transparency required increases when the decision-making algorithm is capable of affecting fundamental values or rights.

When, however, discrimination is traceable to the data and the algorithm replicates the discrimination contained therein due to the statistical method, transparency by design or explicability may not be sufficient. Indeed, the system would not be flawed in the way it works and the discrimination would be a reflection of the inequalities that characterise society. In this case, the corrective could be found not so much in transparency by design, but in the transparency of inputs and outputs. This output control system could also be employed when one is unable to explain how the algorithm arrived at a given decision or when the algorithm has changed as a result of self-learning.

Notwithstanding the inescapable necessity of transparency, there is no shortage of those who point out the difficulty of actually guaranteeing such transparency and an understanding of how the algorithm functions, as well as the risk of opacity that the system maintains insofar as "it is becoming increasingly difficult to explain to non-computer scientists and non-statisticians how a machine learning forecasting model arrives at its results", with the result that "the potential for misunderstandings and even deliberate misrepresentation is vast"⁹⁴:

⁸¹⁸ Thus, on this point, is the idea also proposed G. FIORIGLIO, *La "dittatura" dell' algoritmo: motori di ricerca web e neutralità della indicizzazione. Profili informatico-giuridici*", Bocconi Legal Papers, 3, 2015, 137.

⁸¹⁹ This principle implies that processes should be transparent, the capabilities and purpose of AI systems should be openly communicated, and decisions should, as far as possible, be explained to those directly or indirectly affected by them', and is one of the ethical principles developed by the High Level Expert Group on Artificial Intelligence in its document *Ethical Guidelines for Trustworthy AI*, 14. There are four principles or ethical imperatives: (i) respect for human autonomy, (ii) avoidance of harm, (iii) fairness, and (iv) explicitness.

"even where reverse engineering is possible, understanding of the model remains a matter limited to experts only, to the exclusion of the actual recipients of the 'automated decision'". It is feared that 'algorithmic transparency' is unable to provide the judge, the addressees of the decision and the public with an effective understanding of the process that led to the generation of the digital evidence, and, even more importantly in criminal proceedings, the judgement of its reliability⁸²⁰.

9 *The necessity and clash of penal guarantees*

On closer inspection, the proposal to introduce such instruments in the delicate phase of assessing the dangerousness of an individual, which can affect both the assessment of dangerousness in certain fields and, specifically, in the phase of commensuration of the sanctioning treatment, clashes with and has repercussions on the safeguards provided by criminal law.

For this reason, an attempt will be made here to highlight the main issues concerning them and the possibilities of overcoming them.

9.1 *The difficult balance between presumption of harmlessness and presumption of innocence*

Indeed, "although the two are not synonymous, loss of the presumption of harmlessness has serious implications for the presumption of innocence⁸²¹, which is rightly considered to be a fundamental principle of criminal justice"⁸²².

It is no coincidence that the main questions that lead one to reflect on whether risk assessment and, in particular, on the dangerousness of an individual can be not only presented and proposed but also accepted by the criminal justice system must start from an analysis and reflection on whether the current methods of assessing the risk or dangerousness of an individual are able to provide a reliable but also at the same time complete basis on which to then impose the algorithmic assessment.

⁸²⁰ A. M. MAUGERI, *L'uso di algoritmi predittivi per accertare la pericolosità sociale*, 20 ss.

⁸²¹ On closer inspection, a specific procedural guarantee provided for in Art. 6 para. 2 is the presumption of innocence, according to which "any person charged with a criminal offence shall be presumed innocent until proved guilty according to law." The rule has effects on several levels: as a rule of treatment, in that the accused must be treated formally and substantially as such, until his guilt is legally established by a final judgment; as an evidentiary rule, whereby the burden of proof is distributed between the parties, i.e. the onus is on the prosecution to prove the guilt of the accused; finally, as a rule of judgment, in that if guilt is not proved beyond reasonable doubt, the accused must be acquitted.

⁸²² Please refer also to A. ASHWORTH - M BLAKE, *'The Presumption of Innocence in English Criminal Law'*, in *Criminal Law Review*, 306–17, 1996; V. TADROS, *'Rethinking the Presumption of Innocence'*, in *Criminal Law and Philosophy*, 1, 2007, 193–213.

In other words, in the case of the application of tools such as risk assessment in the assessment of an individual's dangerousness, if the algorithmic assessment of the individual's risk of committing an offence in the future can damage the assessment of the presumption of innocence, one enters an impasse seemingly with no way out.

Obviously, it should be noted that temporally these are different assessments: for while the presumption refers to past offences and facts, the assessment on the probability of committing offences is an assessment that looks to the future. However, as has been pointed out several times, any assessment of risk in itself constitutes a kind of (albeit partial) 'denial of the presumption of innocence', which is difficult to overcome.

Following this directive, in fact, the very fact that an instrument signals or gives as a result a negative score that therefore identifies a particular subject as 'more dangerous' means in part undermining a person and his or her right to be 'presumed innocent in the future'⁸²³.

What is certainly noted is that establishing that in carrying out the assessment and the decision on the sanctioning treatment may in part undermine an assessment that is entirely free from the risk of violating the fundamental guarantees of criminal law, certainly does not pass muster.

Undoubtedly, one of the impasses just underlined could be overcome if, by admitting the algorithmic instrumentation for risk assessment, it could lead to interventions that go beyond the 'best' choice and the 'most personalised and individualised treatment possible' that could thus lead to solutions that are entirely negative for the subject undergoing them. Therefore, as a counterbalance, as will be seen in the following paragraphs, it would be necessary to partially redress this imbalance that is created, to particularly justify this removal and the human decision supported by algorithms.

In reality, as already anticipated, it is noted that one of the limits that criminal law fails to overcome in approaching science and in particular these instruments of prediction are two⁸²⁴: on the one hand, the risk of undermining one of the fundamental guarantees of the criminal justice system, such as the presumption of innocence; however, as has already been pointed out by other voices, this position would not seem entirely convincing since the difficulty is mostly that of admitting predictive judgments of dangerousness of a subject presumed innocent during the trial and one is well aware of the debate on the legitimacy of the precautionary

⁸²³ Indeed, prompted by such concerns, the Swedish Crime Prevention Council has argued that sentencing on the basis of risk assessment is tantamount to making an individual 'serve a sentence for a crime he or she did not commit' and that the practice 'can be compared to convicting an innocent person'. Thus, National Swedish Council for Crime Prevention, *A New Penal System* (1978) cited in Walker, 'Ethical and Other Problems' (n 48), 2.

⁸²⁴ On this point, M. GIALUZ, *Quando l'intelligenza artificiale incontra il diritto penale*, 19.

requirement of prevention of dangerousness; however, once the same are admitted, at that point there is no longer any point in limiting the cognitive instruments of the judge.

Now, the greatest risk derives from the fact that the principle of the presumption of innocence would require that the adjudicating body does not start from any pre-constituted conviction that the defendant did or did not commit the act. Moreover, the rule prescribes that the burden of proof is placed on the prosecution; therefore, this principle may be considered violated if the burden of proof is inevitably reversed, placing it unduly on the defence and not on the prosecution. Consequently, admitting algorithms in criminal proceedings would risk reversing the burden of proof; thus, as already demonstrated, the use of risk assessment tools may unduly influence the perception of the defendant's innocence or guilt.

Undoubtedly, this is not an absolute right since presumptions of guilt clearly operate in any criminal justice system, provided, however, that the arguments are reasonably rebuttable.

Indeed, as already anticipated, the algorithmic opacity flaw makes neither the logical procedure followed by the software, nor at the same time the quality of the data inherent and the specific weight attributed to each of them, unreasonable to the detriment of the defence.

In addition, the data entered cannot be (entirely) neutral, but reflect on different levels the biases of the programmer or of the parties responsible for selecting the data to process the statistical models. Thus, through the indiscriminate collection of the subject's data, a profile of the subject is drawn up, leading him or her to a certain social category.

Consequently, the most problematic aspect, and one that would seem to be without a way out, concerns the fact that it is unthinkable to identify the perpetrator of a given crime solely on the basis of a series of external characteristics; indeed, sentencing must be based exclusively on the facts committed, since 'we cannot impose a penalty if the only tool we have is his or her potential psychological profile, which is not entirely reliable'⁸²⁵.

In conclusion, the elaboration of such profiles, if used to apply a precautionary measure or to modulate a sentence, ultimately end up violating the presumption of innocence, the accused, therefore, will start from a disadvantageous condition since the judge will be conditioned by such external factors and not actual evidence.

9.1.1 *The risk of determinism in decisions*

On closer inspection, as already noted, one of the critical aspects of the application of such instruments is also the risk of incurring a determinism in judicial choices, from which it is

⁸²⁵ J. NIEVA-FENOLL, *Intelligenza artificiale e processo*, 142 ss.

difficult to escape. Indeed, the debate on the applicability of such instruments to criminal justice increasingly raises the risk that even European legal systems, which have long been oriented towards the individualisation of criminal penalties, may suffer a retreat towards deterministic doctrines, also due to the use of such computational models.

The central point and the knot to be unravelled concern, moreover, understanding what kind of use and within what rules and limits the judicial body can make use of such instruments.

In fact, since the human decision-making process is governed by a 'hierarchy of priorities', so the inclusion in the computation model of data referring to the social group to which one belongs, may at the same time mean that the past behaviour of a certain group may affect the fate of the defendant; in such a case, the software would not be able to grasp that innate hierarchy of priorities that governs human action⁸²⁶. Moreover, what value the predictive assessment issued by the algorithm or the risk assessment tools should take on in the judge's decision (in this case on the commensuration of the punishment and thus on the quantification of the sentence) depends on the individual system.

In other words, how much and how far the judge is able to adhere to or detach himself from the predictive assessment issued by the A.I. tool becomes the central question if such an introduction is to be proposed⁸²⁷.

It must in fact be considered that, at its base, the algorithm "is structurally conditioned by the system of values and intentions of those who commission its creation and/or those who elaborate it"⁸²⁸; but to the values identified by the programme - in a democratic and pluralist society - it will be increasingly difficult to conventionally assign a fixed score 'good for all' and therefore objective (as a [deterministic]⁸²⁹ algorithm could do instead)". For this very reason, finally, the algorithm exacerbates the problems and limits of the risk assessment mechanisms themselves and, in particular, of the assessment of dangerousness; the prediction of dangerousness, in particular, presents itself as absolutely problematic because it is not only

⁸²⁶ On this point, QUATTROCOLO S., *Intelligenza artificiale e giustizia: nella cornice della Carta Etica europea, gli spunti per un'urgente discussione tra scienze penali e informatiche*, in *Legislazione penale*, 18 December 2018.

⁸²⁷ It is also no coincidence that even in the Proposal for a Regulation issued at the European level in 2021, the need emerges for full awareness of the so-called automation bias, understood as the possible tendency to uncritically rely on the output of the high-risk AI system, or to overestimate it.

⁸²⁸ S. SIGNORATO, *Giustizia penale e intelligenza artificiale. Considerazioni in tema di algoritmo predittivo*, in *Riv. dir. proc.*, 2020, 614.

⁸²⁹ T. MATHIESEN, *Selective incapacitation revisited*, 455; L. D'AGOSTINO, *Gli algoritmi predittivi per la commisurazione della pena*, 357, who recalls that 'the theory was based on the assumption that professional or trend criminals - responsible for the most serious crimes - can be easily identified from certain known characteristics, such as their personal and criminal history.

inexact, but because it expresses a value judgement that must be adequately reserved to the legal decision-maker.

10 The regulatory framework: the regulatory sources of Artificial Intelligence

As already mentioned, mere human intervention or the increased burden placed on the judge would not help resolve certain issues related to the proposed application of such tools within the sentencing phase. Indeed, human intervention does not help and is not sufficient to correct the errors and discrimination produced by decisions made by artificial intelligence. In order to be able to provide adequate safeguards to individuals who suffer such errors in the criminal field, it is necessary to see and assess what the legal framework is, which, however, has large grey areas to date.

Precisely with regard to the issue of guarantees, and the need to ensure human control, quality, safety, transparency, supranational sources offer significant indications in this regard⁸³⁰. Suffice it to think of the European Ethical Charter on the use of artificial intelligence, or the more recent proposal for an EU regulation establishing harmonised rules on artificial intelligence. Still, Article 11 of Directive 2016/680 on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection and prosecution of criminal offences or the execution of criminal penalties comes into consideration, which establishes a prohibition of decisions based solely on automated processes, if adverse legal effects against the individual may result therefrom.

Indeed, the GDPR provides some safeguards against the effects of automated decisions (albeit not specifically in the criminal field), but shows, at the same time, its limits of protection.

On closer inspection, it can be seen that, although in Italy, as repeatedly stated, there is still neither an embryonic nor a normative attempt to design a regulatory system to regulate the use and utilisation of A.I.

However, this gap, albeit in part, is being addressed at a European level in an attempt to 'govern' the use of such tools.

In doing so, one can see how the attempt is to leave it up to the states to decide whether or not to introduce these new 'subjects', trying to provide them with a regulatory framework so

⁸³⁰ The importance of respecting the principles of transparency, accountability, non-discrimination and accessibility, when the tools in question are used in the context of criminal proceedings, is also reflected in the European Parliament Resolution of 6 October 2021 on artificial intelligence in criminal law and its use by police and judicial authorities in criminal matters (2020/2016).

that they can better govern it, so that they can be users and not victims of these new technologies.

Binding sources certainly include: the so-called 'data protection reform package' of 2016 (consisting of Regulation 2016/679/EU and Directive 2016/680/EU⁸³¹).

Then there is the fundamental rule in truth to be found in Article 11 of the Directive prohibiting decisions based solely on automated processing⁸³² and, in particular, Article 22 of the GDPR.

Still on the soft law side of the sources, on the Greater Europe side instead, there is undoubtedly an extraordinary focus on the growing use of digital tools also in the judiciary. Undoubtedly, the most relevant body of law in this area today is represented by European data protection law. As we have already seen, one of the fundamental factors that has enabled the development of these algorithmic systems, together with the technical development of computing power, is the extraordinary amount of data (personal and non-personal) available today. The regulation of the protection of personal data processing, therefore, was the first regulatory sector to deal with the problem, and the Council of Europe well before the (then) European Community itself, with Convention No. 108 of 1981 on the processing of automated personal data, enforced in Italy by Law No. 98 of 21 February 1989. This was followed by the season opened by Directive 95/46/EC, which Italy implemented with Law no. 675 of 1996. All this complex body of legislation and its subsequent amendments, has recently been reformed and entirely replaced by the new one.

11 New perspectives and positions on artificial intelligence: The EU White Paper

Within the European Union, among other things, the Commission has produced the 'White Paper on AI. A New European Approach, which stipulates that developers of AI systems are already required to comply with European rules on fundamental rights, consumer protection, product safety and liability, without prejudice to the need to implement new rules to deal with the new risks associated with A.I.

⁸³¹ In particular, the latter source represents a kind of *lex specialis* in the area of law enforcement with respect to the Regulation, since it aims to establish minimum rules relating to the protection 'of natural persons with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the protection and prevention of threats to public security'. Thus, see Art. 1(1).

⁸³² On this point, S. SIGNORATO, *Il diritto a decisioni penali non basate esclusivamente su trattamenti automatizzati. Un nuovo diritto derivante dal rispetto della dignità umana*, in *Riv. Dir. Proc.*, 2021, 107 ss.

The Commission proposes a differentiation in terms of discipline, distinguishing between high-risk AI applications and low-risk applications, with the consequence that only the former would be required to comply with certain rules and would have to prove compliance with certain requirements. Of particular interest is the study published in July 2020 in «Artificial Intelligence and Law Enforcement. Impact on fundamental Rights», commissioned by the European Parliament on the impact on fundamental rights, and in particular the right to privacy, of the use of AI tools in policing and criminal justice. The very recent proposal for a Regulation “*Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*”, expresses awareness of the potential of the use of AI tools in 'criminal law enforcement' but, at the same time, considers such tools to be at high risk due to the possible negative impacts on the fundamental rights guaranteed in the Charter, especially when used for individual assessments, by polygraphs and similar tools, to ascertain a person's emotional state, to detect 'deep fakes', to assess the reliability of evidence in criminal proceedings, and, with particular reference to the area under consideration (assessment of social dangerousness), “*for predicting the occurrence or reoccurrence of an actual or potential criminal offence based on profiling of natural persons, or assessing personality traits and characteristics or past criminal behaviour of natural persons or groups, for profiling in the course of detection, investigation or prosecution of criminal offences, as well as for crime analytics regarding natural persons*”.

12 *The Ethic Charter of EU*

One can see how, faced with great changes and upheavals that affect the pillars of criminal justice systems in some cases directly and in others indirectly, Greater Europe finds itself having to take a stand. In fact, since the beginning of the massive spread of such instruments in the North American landscape, Europe has always looked upon them with a sort of reverence, fascination and admiration.

Indeed, the relevance of the issue is also witnessed by the Council of Europe's interest in the growing use of digital tools, governed, in general, by algorithms and more or less sophisticated forms of artificial intelligence, in the field of jurisdiction, as witnessed first by the publication of the study *Algorithms and Human Rights* in March 2018 and then in December of the same year of the 'European Ethical Charter for the Use of Artificial Intelligence in Criminal Justice Systems and Related Environments'. The Charter demands that the development and implementation of artificial intelligence tools and services comply with a number of principles,

including, first and foremost, the principle of respect for fundamental rights, addressed in the first instance to private sector operators, as software developers usually are; the Charter thus recalls both the European Convention on Human Rights and the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (adopted in Strasbourg on 28 January 1981 and made enforceable in Italy by L. 21 February 1989 no. 98), which establish some principles considered relevant in the activity of creating software, as a tool, inter alia, to support jurisdictional decision-making: the principle of legality; the right of access to jurisdiction; the right to a fair trial, in its essential articulations of cross-examination and equality of arms; the independence of the judiciary and, in particular, of judges in the exercise of decisional power.

Indeed, in the face of the continuous spread of these instruments, the debate has also started at the European level, while maintaining the non-interference in the 'Chinese walls' of each legal system⁸³³. The debate focused first of all on the major points of criticism and friction encountered in the first overseas applications. Indeed, arriving 'by seconds' in a debate that clearly involves jurisdictions with different bases and institutions, can nonetheless be useful in providing a picture of what problems have emerged since the first applications.

In fact, the question has been raised as to how judicial systems will be able, in the near future, to cope with these technological developments, trying to control them and frame their use in a clear manner with respect to fundamental rights and guarantees.

In the context of Wider Europe⁸³⁴, attention to the use of these tools has grown over time. An important step was in fact the adoption of an important softlaw instrument, the European Ethics Charter⁸³⁵ for the Use of Artificial Intelligence in Justice, which was adopted by the Commission on the Efficiency of Justice (CEPEJ)⁸³⁶.

⁸³³ V. MANES, *Il Giudice nel labirinto*, Rome, 2012.

⁸³⁴ In March 2018, a study was published in Algorithms and Human Rights, which formed an important basis for the adoption of the European Ethics Charter in December of the same year. Regarding the study referred to, please refer to Algorithms and Human Rights - Study on the human rights dimension of automated data processing techniques and possible regulatory implications.

⁸³⁵ In fact, the document is addressed to «public and private stakeholders responsible for the design and deployment of artificial intelligence tools and services that involve the processing of judicial decisions and data», nonché ai «public decision-makers in charge of the legislative or regulatory framework, of the development, audit or use of such tools and services»; on this point M. GIALUZ, *Quando l'intelligenza artificiale incontra il diritto penale*, 12. Regarding the document, please refer to European Commission for the efficiency of justice (CEPEJ), *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*, 3rd – 4th December, 2018, 5.

⁸³⁶ The document was drafted by the CEPEJ, the European Commission for the Efficiency of Justice, established in 2002 at the initiative of the Committee of Ministers of the Council of Europe, with the aim of monitoring and measuring the quality of the justice systems of member countries. For a first reading commentary QUATTROCCOLO S., *Intelligenza artificiale e giustizia: nella cornice della Carta Etica europea, gli spunti per un'urgente discussione tra scienze penali e informatiche*, in *Legislazione penale*, 18 December 2018.

The peculiarity of this document is its structure and the choice of identifying five main points (the five principles) that underpin and act as a 'beacon' in the use or possible introduction of these instruments within European legal systems.

An attempt will therefore be made to describe, albeit briefly, the content of the various principles.

Firstly, when artificial intelligence tools are used as an aid in trials, it must be ensured that they do not violate the right of access to a judge and the right to a fair trial (with all that is meant by this: respect for the equality of arms, cross-examination).

Secondly, then, the canon of non-discrimination is enshrined, since the capacity and characteristic of these instruments to use and process data, in some cases providing discriminatory results, is brought into play. This second principle includes the need for those who use such tools, whether public or private entities, to ensure that they do not reproduce ex novo or aggravate such discrimination (especially when it comes to tools that process or take into consideration personal data of individuals concerning family characteristics, economic or social background, political opinions, or data concerning geographical origin and provenance, sexual orientation, etc.).

Indeed, while admitting the risk and the possibility that such discrimination will inevitably occur, corrective measures must nevertheless be considered in an attempt to limit or eliminate such risks altogether.

Next, the third principle indicated by the Charter relates to quality and security: it is recommended to use data (and all data derived from judicial decisions) that clearly come from certified sources; furthermore, the process must be traceable and the models and algorithms created must be able to be stored and executed in secure environments, so that the integrity of the system can be guaranteed.

As a continuation, the fourth essential canon, as transparency is prescribed, is linked to the requirements of impartiality and fairness: indeed, accessibility to the algorithmic process, absence of bias and intellectual integrity must also prevail over the requirements of intellectual property protection⁸³⁷.

⁸³⁷ These values can be ensured first and foremost by complete technical transparency (of the source code and documentation), which, however, does not appear to be sufficient in itself: it has been correctly noted that, "even where reverse engineering is possible, the understanding of the model remains a matter limited to experts only, to the exclusion of the actual recipients of the 'automated decision'". Thus, on this point and in these words, S. QUATTROCCOLO, *Intelligenza artificiale e giustizia*, 8.

On closer inspection, the last canon to be taken into consideration is also defined as under user control, by virtue of which a prescriptive approach to the use of artificial intelligence is also ruled out and it must also be ensured that users, at the same time, act as informed subjects who are in full control of their choices.

As far as this principle is concerned, in particular, since the user can be both the subject who sees his or her acts used within the process for the purposes for which the tools are set up and introduced, but at the same time the same can also be used by the legal practitioner. Therefore, looking at this double facet or double medal to which it is addressed, one can see how this principle, in the first place, translates into the possibility of re-examining the decisions and data used to produce a result and still not being bound to the solution provided by the I.A. tool, or at least not only to it. In fact, it is precisely on this point that the greatest difficulty arises, since it is precisely the establishment of parameters or a rough indicator that establishes a priori how far a judge can or must deviate from an algorithm that is one of the thorniest issues to be resolved⁸³⁸. Therefore, connected to this is the need for the subject of the decision to be able to review and have full access to the decision and the data used to produce a result.

After all, the fundamental point reaffirmed by the Charter is the right to be informed of the different options available and the right to legal advice and to have access to the court, according to the provision protecting this principle under Article 6 ECHR⁸³⁹.

13 The Council of Europe's position on automated decisions with profiling

On closer inspection, the most complex and sophisticated automated decisions are usually made precisely because of the results and insights obtained from profiling. Indeed, the more precise and exact the profiling process, the more accurate the decision that will be made. The

⁸³⁸ The 'European Ethical Charter for the Use of Artificial Intelligence in Criminal Justice Systems and Related Environments' of 2018⁷³, cited above, pointed out in relation to criminal proceedings that 'even if they are not specifically designed to be discriminatory, the use of AI-based algorithms [. has shown the risk of fostering the resurgence of deterministic theories to the detriment of the individualisation theories of punishment' (p. 48); thus, the prohibition of discrimination against groups or individuals in the use of computational processes used in judicial proceedings is affirmed. In the proposed Regulation (Artificial Intelligence Act) 21 April 2021, it is pointed out that if the AI system is not trained with high quality data, does not meet the necessary accuracy or robustness requirements, or is not adequately designed and tested before it is placed on the market or otherwise put into service, it may identify individuals in a discriminatory or otherwise unfair manner.

⁸³⁹ Moreover, in the first appendix to the Charter - containing a Study on the use of AI in judicial systems - on the one hand, the critical issues linked to possible discriminatory effects of predictive liability tools are reiterated and, on the other, respect for the principle of equality of arms, the presumption of innocence and the need for the interested party to be able to challenge the scientific validity of the algorithm and the weight given to the various data is emphasised: In this perspective, the key is the right of access to the judge, which is also grounded in the principles of personal data protection. See, European ethical Charter, Appendix I, In-depth study on the use of AI in judicial systems, notably AI applications processing judicial decisions and data, § 138.

added value, therefore, of this particular type of decision is represented by the very activity of profiling itself, which provides a sort of complete and detailed picture of the individual.

One of the debates on the subject concerned precisely the attempt to provide a definition of profiling. In this regard, it was the Council of Europe that in its 2013⁸⁴⁰ Recommendation clarified and outlined the lines of the concept of profiling, which can include within it three different stages:

1. The data collection phase. This first phase, which forms an integral part of the decision-making process, is entirely useful in order to obtain profiling (and a decision) that is as effective as possible.

2. Automated analysis to identify different correlations. This phase is also called learning and normally works as follows:

- Learning begins and starts with selected information that contains patterns or similarities;
- The patterns are then identified, and it is at this stage of identification that machine learning algorithms perform particularly well;
- A model is generated that is able to recognise patterns when new data is processed.

3. Application of a model to an individual or set of individuals to identify present or future common behavioural characteristics. In turn, this phase then includes several steps whereby the model operates following a pathway divided into several stages: searching for new (and so far or not considered) data; deciding which profile those data are closest to; producing a result.

14 Regulatory limits to artificial intelligence: ethical and legal barriers?

After all, there is a legal framework, albeit not a uniform one, for this discipline. In what follows and in the concluding part of this paper, an attempt will be made to provide an overview of the most relevant supranational and national regulations on Artificial Intelligence and, more specifically, of the consequences arising therefrom on the unlawful processing of personal data.

Finally, another level that needs to be questioned relates to the nature of human decision-making and the consequent risk of its distortion. Indeed, human decision-making activity is complex and cannot be summarised in merely mathematical language. In fact, the judge, through inductive and deductive reasoning, performs an analysis and synthesis of the elements of the judgement and also arrives at the judgement on the basis of his or her own knowledge, experience, studies and analytical skills. Indeed, the concept and activity of 'judging' represents

⁸⁴⁰ CM/Rec (2010) of the Committee of Ministers to Member States on the Protection of Individuals with Regard to automatic Processing of Personal Data in the Context of Profiling.

a 'combination of knowledge, formulation and verification of hypotheses and also the interference of human emotions that, in some cases, succeed in "adapting justice to the concrete case"⁸⁴¹.

It seems useful to dwell on the nature and characteristics of the judge's decision-making process; in fact, the two main models of 'judging' developed by science refer to the formalist and realist conceptions.

According to the former, judges apply the law to the facts of the case in a logical and mechanical manner: in this sense, the judge behaves according to a sort of 'giant syllogism machine'⁸⁴². On the contrary, with the second conception, judges follow an intuitive process to come to conclusions that only later rationalise with reasoned reasoning.

On closer inspection, one cannot limit oneself to the description of only these two models as it does not appear satisfactory in describing the cognitive process underlying the judgement. In fact, it is necessary to integrate the output of both models in order to obtain a more complete picture of the human being's decision-making mechanisms. In fact, it cannot but be considered that intuitions play a key role in the first part of the process, but then more weighed and complex, less 'automatic' considerations intervene to act as a 'corrective'.

The nature of the process is therefore twofold: inductive, i.e. spontaneous, fast and automatic at first, deductive requiring subsequent mental effort; mental processes at this stage are so-called 'deliberate rule-governed effortive, and slow'. Consequently, the first moment proposes intuitive answers to legal problems as they arise; only in a second moment are the qualities of the considerations assessed, which can be corrected, reconsidered or confirmed.

It is noticeable that judges have shown a greater tendency towards the intuitive approach, also by virtue of the stimuli they receive at the judgment stage. Indeed, it has often been observed that judges are vulnerable to hindsight bias: in fact, judges, by evaluating facts after they have happened, run the risk of overestimating the predictability of certain events.

In conclusion, one can see how this incessant clash and confrontation between technologies and human activities characterises today's societies and, even more specifically, it is a dialogue that can neither be avoided nor set aside.

There would undoubtedly need to be a self-reflection on the concept of judging, on the role that is deferred to human beings and on how difficult it is to accept the element of risk not only

⁸⁴¹ See, J.NIEVA-FENOLL, *Intelligenza artificiale e processo*, 46.

⁸⁴² C. GUTHRIE – J. J. RACHLINSKI - A. J. WISTRICH, *Blinking on the Bench: How Judges Decide Cases*, Cornell Law Faculty Publications, Paper 917, 2007, 2.

on a possible miscarriage of justice, but also on the choice of a certain amount of punishment that is free of prejudice and intuitive elements.

In a future and utopian scenario, one would have to imagine a reality in which the reproach one decides to impose on a subject for whom the decision on guilt has already been made corresponds to a criterion of 'justness'. Clearly, this part of the idea is not feasible and would call into play, at the same time, other elements, including a reflection on whether it is right to delegate the choice on personal freedom to another individual, questioning the legitimacy and at the same time the necessity of decisions that are detached from the concrete fact and require a forward-looking effort.

In spite of this, it is necessary to reflect on opportunities in order to understand whether they can 'improve' a decision-making moment that presents many elements of difficulty. However, keeping to a still embryonic idea and perspective, one could imagine a future that looks first and foremost to a dialogue between public authorities, algorithmic companies and the judicial system. The involvement of the public sector in the provision of accessible and reliable tools, with a broad contribution of jurists and experts from different disciplines, could lead to new and hopefully more accurate ways of taking decisions involving predictive reasoning, even at the investigation stage. And, so much so, without, of course, suppressing the intuitive moment of decision-making, which, as recent studies have also shown, has a positive impact on the deliberative process. Only the judge can, in the end, in adapting the statistical and experiential rules to the peculiarities of the concrete case, adequately fulfil the adoption of decisions that have a strong impact on fundamental rights. On the other hand, where an assessment of a psycho-criminological nature is taken into consideration, or where it is a question of sifting through the programming rules of a certain algorithm, judicial scrutiny according to the criteria of technical-scientific proof, as well as the performance of an adequate cross-examination, also by means of the contributions of the party's technical consultants, appear not to be renounced.

In conclusion, the intrinsic limitation of the processing capabilities of the human mind, as well as of the time and resources available, and the hope for greater accessibility of the cognitive distortions interfering with the evaluations at issue - through the imposition of transparency requirements in the selection of data and in the programming of the algorithm - seem to make the prospect of the aid of technological tools a challenge to be tackled with a positive orientation, provided that one proceeds in the provision and strengthening of guarantees, according to the cues already offered by supranational sources.

14.1 *The General Data Protection Regulation*

It should be stated at the outset that the General Data Protection Regulation (GDPR) does not directly address the subject of artificial intelligence, but contains relevant rules on information, profiling and automatic decision-making. In particular, this regulation intervenes on the topic of most relevance here, namely that of profiling. Its peculiarity lies in the fact that it does not merely prohibit profiling (which receives its own definition in Article 4), but, on the contrary, requires it to have a legal basis (Art. 6) and to be based on appropriate mathematical or statistical procedures (in Recital 71), in accordance then with the guidelines indicated by the European Data Protection Committee (Recital 72).

In particular, in order to determine the lawfulness of profiling, even in the presence of a legal basis, the guidelines drawn up by the Article 29 Working Party require the following factors to be taken into account: the level of detail and completeness of the profile (whether it only describes partial aspects of the data subject or reconstructs a more complete picture); the impact of profiling on the data subject; and the security measures to ensure fairness, non-discrimination and accuracy of the profiling process.

In addition, the GDPR lays down a general prohibition on subjecting individuals to fully automated decision-making processes. This prohibition, however, is also countered by several exceptions. Indeed, automated decisions are permitted if their use: is necessary for the conclusion or performance of a contract between the data subject and the data controller; is authorised by a law or regulation; and is based on the explicit consent of the data subject.

Moreover, in the event of profiling and automated decision-making processes, the GDPR guarantees the subject concerned the right to be informed⁸⁴³.

Furthermore, in the case of profiling and automated decision-making processes, the GDPR guarantees the data subject the right to be informed.

In particular, the data controller is also obliged to inform data subjects about: the methods and purposes of profiling; its logic; and its consequences.

The right to information is also accompanied by the right to object to profiling (Article 21), to request the deletion of one's own data and profile (Article 17) and to contest automated decisions (Article 22(3)).

There is another provision that relates to and concerns, more specifically, the profiling and processing of data, with a view to applications that take place in a fair manner. Indeed, Art. 4 of the Regulation [T]he 'profiling' [...] consists of any form of automated processing of

⁸⁴³ This is found in Articles 13, 14 and 22 and Recital 71 of the GDPR.

personal data that evaluates personal aspects relating to a natural person, in particular in order to analyse or predict matters such as the data subject's work performance, economic situation, health, personal preferences or interests, reliability or behaviour, geographical location or travel patterns, where this produces legal effects concerning him or her or significantly affecting him or her'.

14.2 Automated data processing: a step forward to the GDPR with Legislative Decree No. 51 of 2018

The General Data Protection Regulation (GDPR) was introduced into Italian law to comply with European legislation on the processing of personal data. It was then transposed into our legal system in 2018⁸⁴⁴.

The legislative decree regulating the wholly or partially automated processing of personal data for the purposes of preventing and suppressing crimes, executing criminal sanctions, and safeguarding against threats to public safety is relevant because it extends the precautions contained within the GDPR to criminal law.

In particular, Article 8 of the Decree would seem to reproduce almost the entirety of Article 22 of the GDPR but, unlike the latter, it enshrines the absolute prohibition of decisions based solely on automated processing in criminal matters⁸⁴⁵.

It also specifies that the decisions referred to in point 1 of the same provision cannot be based on special categories of personal data (religion, race, sex), unless appropriate measures are in place to safeguard and protect the rights, freedoms and legitimate interests of the data subject. The same decree also provided for two new criminal offences: the offence of profiling, aimed precisely at discrimination, and that of unlawful processing of sensitive data.

The GDPR does not directly address the issue of A.I., but nevertheless contains relevant rules on information, profiling and automatic decision-making 136 The GDPR does not prohibit profiling (as defined in Art. 4, no. 1), but requires it to have a legal basis (Art. 6), and to be based on 'appropriate mathematical or statistical procedures' (Recital 71), in accordance with the guidelines set out by the European Data Protection Board (Recital 72). In order to

⁸⁴⁴ In particular, it was transposed within our legal system by Legislative Decree No. 51 of 18 May 2018 implementing EU Directive 680/2016150 on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection and prosecution of criminal offences or the execution of criminal penalties.

⁸⁴⁵ As stated in the same provision, in fact, 1. Decisions based solely on automated processing, including profiling, which produce adverse effects with regard to the data subject shall be prohibited, unless authorised by EU law or specific legal provisions. 2. The legal provisions must provide adequate safeguards for the rights and freedoms of the data subject. In any case, the right to obtain human intervention by the data controller shall be guaranteed'.

determine the lawfulness of profiling, even in the presence of a legal basis, the guidelines developed by the Article 29 Working Party (Opinion 2016/679) require the following factors to be taken into account: - the level of detail and comprehensiveness of the profile (whether it only describes partial aspects of the data subject, or reconstructs a more complete picture of the data subject); - the impact of the profiling on the data subject; and - the security measures to ensure fairness, non-discrimination and accuracy in the profiling process. The GDPR, on the other hand, enshrines a general prohibition on subjecting individuals to fully automated decision-making processes, including profiling, capable of producing legal effects or significantly affecting the data subject (Art. 22).¹³⁷ However, there are broad exceptions to this prohibition. Automated decisions are permitted if their use: (i) is necessary for the conclusion or performance of a contract between the data subject and the data controller, (ii) is authorised by a law or regulation, or (iii) is based on the data subject's explicit consent. Furthermore, in the case of profiling and automated decision-making processes, the GDPR guarantees the data subject the right to be informed (Art. 13, 14, 22 and Recital 71 of the GDPR). In particular, the data controller has an obligation to inform data subjects about (i) the modalities and purposes of the profiling, (ii) its logic and (iii) its consequences. The right to information is accompanied by the right to object to profiling (Art. 21), to request the deletion of one's own data and profile (Art. 17), and to contest automated decisions (Art. 22(3)).

14.3 EU Directive 680/2016 on the processing of personal data for the prevention, investigation, detection and prosecution of criminal offences or the execution of criminal penalties

Directive 2016/680 of the European Parliament and of the Council of Europe stands as a bulwark for the protection of natural persons with particular reference to the processing of personal data by the competent authorities for the purposes of prevention, investigation, detection and prosecution of criminal offences or the execution of criminal penalties, as well as the free movement of such data; at the same time, it introduces the regulation of the protection of natural persons with reference to the processing of data by the authorities for the purposes of prevention, investigation and prosecution of networks.

On closer inspection, the Directive entered into force in May 2016 and has only been implemented since May 2018⁸⁴⁶.

⁸⁴⁶ In Italy, the transposition took place with Legislative Decree No. 51 of 18 May 2018.

Besides regulating the processing of data by law enforcement authorities, it also aims to strengthen judicial cooperation in criminal and police matters, and thus to achieve a more efficient exchange of information between authorities.

The new legislation replaces that in Titles I and II of Part II of the Privacy Code, dedicated to the judiciary and police processing.

The scope of application is limited to wholly or partially automated processing of personal data and to non-automated processing of personal data contained in or intended to be contained in a file, carried out by the competent public authorities for the prevention, investigation, detection and prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and prevention of threats to public security and public powers for the same purposes as above.

However, the decree does not apply to data processing carried out in the course of activities concerning national security and carried out by EU bodies or agencies.

Likewise, it does not apply in the chaos of data processing relating to other tasks conferred on the competent authorities and not necessarily carried out for the prevention, investigation, detection or prosecution of criminal offences (in that chaos, the European Data Protection Regulation applies).

The directive applies to personal data exchanged between member states but also to the processing of personal data carried out at national level for police and judicial purposes.

This is intended to facilitate cooperation between police or judicial authorities.

Furthermore, data collected by police authorities must fulfil the following conditions: they must be processed lawfully and fairly; collected for specific and explicit purposes; adequate and relevant and not excessive in relation to the purposes for which they were collected; accurate and up-to-date; kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which they were collected; processed in such a way as to ensure their security and to prevent unauthorised processing and the loss or destruction of the data.

Furthermore, the purpose principle is subject to limitations in the sense that personal data collected by competent authorities for law enforcement or public security purposes may be used for different purposes provided that such use is authorised by Union or Member State law. However, as a Member State, the GDPR applies.

15 *The proposal for a future regulation on I.A. de iure condendo perspectives. Progress towards regulation? The proposed Regulation on A.I.*

It should be noted that it was precisely in lamenting the lack of a specific ad hoc regulation on A.I. that the European Commission recently presented a proposal for a Union regulation that would like to create a single body of legislation regulating A.I. in its entirety (in essence, creating a special A.I. law).

Recently, the definitional and legal problem on the definition of Artificial Intelligence has been addressed and taken into consideration by the recent proposal for a Regulation on AI. Indeed, the idea behind it was to create a single discipline that has been advocated and requested by several states.

This is a real AI law that was recently presented by the European Commission.

The regulation, in particular, also had to provide its own definition of A.I. in order to then be able to define the scope of application⁸⁴⁷.

The most difficult element was to find a definition of A.I. that could be both complete and precise; the difficulty stems from the fact that it is possible to understand A.I. as a single scientific and technological discipline, but within it it encompasses a very wide range of methods and techniques that are applied to a very broad and diversified set of scientific, technological and industrial objectives. The regulation and proposal have therefore also embraced an all-encompassing idea of definition, which must be read in a teleological key so as to encompass as far as possible all and only those systems that present risks and opportunities of typical intelligent applications.

It is therefore no coincidence that as a direct consequence of this choice. the most significant rules of the Regulation apply only to systems that the Regulation itself classifies as 'high-risk systems'. Indeed, what is most relevant for the purposes of applying the Regulation is not so much the qualification of "A.I. systems", but rather the fact that the system in question must fall within one of the categories of high-risk systems⁸⁴⁸.

⁸⁴⁷ In Article 3 of the regulation, a definition of an AI system is proposed, i.e. software developed with one or more of the techniques and approaches listed in Annex I that can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations or decisions that influence the environments with which it interacts. Annex I of the same Regulation distinguishes three technologies that characterise AI (the use of which indeed seems sufficient to attribute the nature of 'AI system' to software based on them): a) machine learning approaches, including supervised learning, unsupervised learning and reinforcement learning, using a wide range of methods, including deep learning; b) logic-based and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inferential and deductive engines, (symbolic) reasoning and systems.

⁸⁴⁸ See, G. SARTOR, *L'intelligenza artificiale e il diritto*, 9.

As already mentioned, it was decided to adopt a very broad concept of A.I. in the Regulation, which extends to all applications adopting "machine learning approaches", logic-based and knowledge-based approaches and "statistical approaches, Bayesian estimation, search and optimisation methods".

Indeed, the Regulation adopts at the same time a 'risk-based' perspective: in other words, the main objective it pursues is not to provide individuals with new legal remedies against the prejudices deriving from prohibited uses of A.I., but rather to prevent those prejudices in this way: on the one hand by regulating the development, distribution and use of A.I. systems; on the other by setting up standards and controls, the definition and verification of which is entrusted to private and public structures. It would almost seem that the regulation has decided to follow a sort of 'preventive' line (as is done in the case of foodstuffs).

In fact, in an attempt to describe the structure of the Regulation, it can be immediately anticipated that the discipline envisaged by the Regulation is based on a classification of A.I. applications into different risk categories⁸⁴⁹. The Regulation, therefore, after clearly specifying (in Art. 5) those which are the prohibited uses of A.I., assumes on the other hand that all the others are lawful, if they do not violate other legal norms⁸⁵⁰.

Moreover, Article 53 of the Regulation introduces some specific information obligations to protect the subjects, to ensure that they are not 'deceived' or 'manipulated' by A.I. systems⁸⁵¹.

⁸⁴⁹ Applications that pose an unacceptable risk to individuals and society are prohibited by Art. 5 of the regulation. These are the following categories - systems that use 'subliminal techniques that act without the knowledge of a person in order to materially distort that person's behaviour in a way that causes or is likely to cause that person or another person physical or psychological harm'; - systems that exploit 'the vulnerabilities of a specific group of persons, due to age or physical or mental disability, in order to materially distort the behaviour of a person belonging to that group in a way that causes or is likely to cause that person or another person physical or psychological harm'; - systems used by public administrations 'for the purpose of assessing or ranking the trustworthiness of natural persons' by assigning a 'social score' that may lead to unfavourable treatment [...].

⁸⁵⁰ There was no lack of voices calling for an extension of the prohibitions in Article 5. In particular, the following aspects were criticised: the limitation of the prohibitions in points (1) and (2) to physical or psychological harm only, with the exclusion of economic and moral damage; the limitation of the prohibition in point (3) to public administration activities; and the exceptions to the prohibition of biometric identification systems in point (4), which allow a surveillance infrastructure to be put in place, potentially usable for other purposes. These criticisms were countered by the need for the legislator to intervene cautiously with respect to A.I., a new and rapidly evolving sector.

⁸⁵¹ For more details, please read the provision. In particular, persons interacting with an A.I. system must be informed that their interlocutor is an artificial system rather than a human being, unless this is evident from the circumstances and context of use; persons exposed to an emotion recognition system or a biometric categorisation system must be informed about the functioning of such a system; it must be disclosed that a piece of content has been artificially generated or manipulated, if that content may appear 'falsely authentic or truthful'.

In fact, it should be noted that the most important provisions of the regulation are those aimed at applications reported as 'higher risk', which are lawful provided they comply with the requirements concerning them and which are contained in Title III⁸⁵².

⁸⁵² These are the following systems (Art. 6): AI systems that are governed by regulations on high-risk products (the harmonisation regulations listed in Annex II) and for which a third-party conformity assessment is envisaged; AI systems listed in Annex III of the regulation (Biometric identification and categorisation; Management and operation of critical infrastructure; Education and vocational training; Access to and use of essential public and private services and facilities; Law enforcement; Management of migration, asylum and border control; Administration of justice and democratic processes). Whoever provides or employs a high-risk system must ensure that the system complies with the following requirements: establish, implement, document and maintain a risk management system (Art. 9); adopt quality criteria for data used in training (Art. 10); - prepare technical documentation demonstrating compliance with the applicable requirements (Art. 11); allow automatic logging of events ("logs") during their operation (Art. 12); - ensure that their operation is sufficiently transparent to allow users to interpret the system's output and use it appropriately (Art. 13); - be effectively supervised by natural persons (Art. 14); - achieve an adequate level of accuracy, robustness and cybersecurity (Art. 15); The provisions on high-risk systems are also the subject of wide-ranging debate: some believe that the requirements listed are not sufficient to ensure the reliability of AI systems; others, on the other hand, consider them to be excessively restrictive and such as to compromise the competitiveness of European industry. First of all, it is stipulated that the assessment of IA systems is to be entrusted to 'conformity assessment bodies,' whose capabilities are to be assessed in advance by 'notifying authorities' designated by the member states (Art. 30 and 31). These bodies will issue certificates of conformity allowing the use of high-risk systems. Secondly, each Member State will have to establish or designate competent national authorities to ensure the application and implementation of the regulation (Art. 59). A National Supervisory Authority will be appointed at the head of these authorities, with the functions of notification authority and market surveillance authority. The National Supervisory Authorities and the European Data Protection Supervisor form the European Data Protection Board, which provides assistance to the Commission and the National Supervisory Authorities and coordinates their activities.

Conclusions

The general uncertainty that characterizes an era in which technology is constantly making giant strides has led man to increasingly resort to tools that can in some way reduce and at the same time overcome it. In fact, one cannot fail to realize that we have entered a new era of so-called '*technohumanism*'; therefore, it is possible to rethink the relationship between man and machine, in which cognitive performance can be merely enhanced by artificial intelligence, without, however, disregarding or forgetting the inconsistencies and contradictions of the human decision-making process, and without going so far as to replace it⁸⁵³.

In this field, algorithms have positioned themselves as a kind of '*pharmakon*', capable of counteracting the fallibility of human evaluations.

On closer inspection, precisely in the society of 'performance' in which the imperative is efficiency 'at all costs', algorithms have found fertile ground for expansion into every field (in fact, they are present in almost every day-to-day sphere).

In the course of the discussion of this thesis, an attempt has been made to demonstrate the ambivalent character of these tools; indeed, it would be reductive and perhaps oversimplistic to take a clear-cut position in opposition to them, also because the benefits they have brought in various fields cannot be disregarded.

Clearly the discourse, when analyzing the issue by shifting it to the level of criminal justice where algorithms have also gradually made their timid entry, becomes and more delicate fraught with pitfalls.

During the course of the research, I was realized that, even if one had started out in the first fascinating researches in this field, the importance and necessity of properly understanding how the use of algorithms in the field of justice should not be demonized or blindly dismissed per se, but needs careful study and discernment of the many possible applications. The synergy between man and machine has always fascinated human beings and may constitute a valid support to decision-making activities; however, even when envisaging such an eventuality, albeit on a theoretical level, it would be necessary to set limits and uniform guidelines at a national and even supranational level.

Starting precisely from a comparative analysis with the US legal system, it was possible to analyze the current state of things and see how the use of risk assessment tools in the sentencing

⁸⁵³ See A. PUNZI, *Judge in the Machine: e se fossero le macchine a restituirci l'umanità del giudicare?*, in A. Carleo (ed), *Decisione robotica*, Bononia, 2019.

phase most often results in a surreptitious circumvention of fair trial guarantees⁸⁵⁴. For this reason, it was deemed necessary to investigate this field.

The initial reflection in this paper started from a fundamental element of awareness: that is, from the fact that, in today's criminal trial, there are numerous moments in which the judicial authority is called upon to carry out evaluations of a prognostic-predictive type; there are therefore numerous cases that burden the decision-maker with an extremely complex prognosis (therefore characterised by elements of uncertainty).

This research study has therefore sought to provide an answer to macro-questions that primarily intersect a question - albeit a parallel one - that governs the meaning and basis of criminal justice: the concept of judgement. This contribution, in particular turns its attention primarily to this last phase, in which predictive algorithms act as a guiding tool - optional, but in some cases mandatory (in the case of the US legal system) - in the deliberation of the conviction. In the course of the development of the paper, it was premised from the outset that this capacity for judgement calls into play various sciences and studies pertaining to different human capacities, who in his analysis and rational development finds himself having to make certain choices.

The algorithmic assessment of dangerousness leaves open numerous questions on the respect of the defendant's safeguards, on the reviewability of the final result, on the scientific falsifiability of the software, on the reliability of the inputs and output, on the residual duty of motivation on the part of the judge, on the discriminatory effects arising from empirical *generalizations* and the social and economic conditioning factors processed by the algorithm.

More specifically, it was decided to focus on a particular type of choice that nevertheless requires the judge to go one step further: to look into the future and, on the basis of objective, subjective and factual elements, decide what is the best punishment for a given offender.

Having regard in the fourth chapter to the judge's activity in the commensuration of punishment, it was considered that the entry of predictive software could make possible a more accurate prognosis of the offender's capacity to commit offences and the assessment of the individual parameters indicated in the second paragraph of Article 133 of the criminal code. The aim of making the judicial choice more 'complete' may be pursued by means suitable for

⁸⁵⁴ As already analysed in more detail in Chapters 4 and 5, deregulation, unscrutinable results, the implicit biases contained in algorithmic software and the risk of delegating the decision to the algorithm (deemed neutral and objective by virtue of the automation bias) risk undermining the aforementioned guarantees in the absence of external control.

preserving the guarantees typical of the criminal trial. Particular caution must be exercised in selecting only those predictive tools that compare the results of the individual examination with statistics relating to similar criminological profiles (so-called mixed-type assessment tools).

In doing so, as we have seen, the choice of the best sanctioning treatment looks both to the past and to the future. On the one hand, in fact, it cannot but look at the fact committed and modulate itself so that therefore the reprimand is perceived and respected, in proportion to the fact committed; but, at the same time, it also looks to the future since it must necessarily question itself as to what the subject's risk of relapsing into error or instead of being re-educated might be.

What has certainly been noted during the development of this discussion is that from the reflection on the moment of the choice of sanctioning treatment (which implies within it an assessment that looks to the future) many questions emerge. On the one hand, in fact, there are still unresolved aspects on the legislative side; in fact, the vagueness of the formulas identifying the object of the judgement; the lack of precise indications as to the framework within which the prediction is referred to and required to be made; the uncertainties as to the standard of ascertainment and also the existence of limits placed on the judge's acquisition of the information that is indispensable for establishing the evolution of the offender's behavior.

There is also a serious lack of empirical data and guidance on how to assess the risk of reoffending. This penal system, thus hinged, clearly clashes with a certain reluctance and diffidence in resorting to and calling into play other human sciences, in addition to clinical-statistical instruments, in order to be able to assess the prognostic judgement, following canons of sufficiently corroborated controllability and verifiability.

In reality, what has emerged and been noted is that it is precisely the contribution of scientific knowledge that seems to lose its indispensable character when it comes to formulating prognostic judgement. This is even more true at the moment of choosing a commensurate penalty.

Indeed, the analysis carried out also shows how the shortcomings that afflict the role of prognosis in the penal system cannot in any way justify a renunciation of the judge's room for discretion, which in fact remains indispensable.

The result to which this work aspires is to identify a line and draw a direction on what are the current applications of other systems that were until recently foreign to the penal system. In addition, the aim is to try to look at the delicate moment of the choice of punitive treatment and to see how the same is rather weak and the prognostic judgement is central. It is noted how this type of judgement assumes a fundamental importance and centrality in the punitive system,

which is not always considered. It is central above all because such evaluations, which cannot stop at being mere conjecture or intuition on the part of the adjudicating body, make the actual execution of the penalty (or security measure) threatened by the legislature or then concretely identified by the judge depend on it.

Indeed, the legislative provision of the prognostic judgments entrusted to the judge is a choice that is constitutionally per se obligatory: in fact, only by casting a glance into the future can one imagine that re-educative pathway that Article 27(3) places in the foreground in order to enunciate the purposes of punishment.

Indeed, it is only by means of the forecast, even if inevitably uncertain, of the offender's behaviour, that the principle of individualization of the punitive treatment and of the least sacrifice necessary can be implemented and realised; in fact, the *extrema ratio* is to be found both in the *an* and in the *quantum* of the punitive (or even para-punitive) response and is realised through the prognostic judgement.

On closer inspection, this is an essential step in adapting the response to the crime to the individual. In fact, the aims of special prevention change the perspective of criminal law, since it must necessarily turn towards the future, i.e. towards that probability of improvement that is the essence of human nature. If, on the other hand, one places oneself in a different perspective and moves from the direction marked by the more retributive ideal, i.e. the injury of the past to be compensated for, towards the course indicated by special prevention, i.e. the future to be changed and the offender to be ensured re-educative treatment.

Therefore, on the prevention of recidivism and the identification of the special-preventive effects of the response to the offence, the effective protection of legal goods depends. These are objectives that can only be achieved through the contribution of prognoses.

Of prognoses that therefore go beyond mere human intuition and exclusive reference to common culture populated by maxims of experience that are not always well-founded.

The investigation of this study started precisely from the primary consideration of whether the reality of prognostic judgements can actually be modified to conform to the requirements mentioned.

For example, with particular reference to the determination of the perimeter of the assessment of the risk of recidivism, we have already seen how the judge seems to be instinctively led to delimit the object of the prognosis. Indeed, he would seem to make his judgement around narrower categories of offences (e.g. in determining whether the subject will commit offences of the same nature).

In other words, an initial limitation of the subject of the prognostic judgment would seem appropriate: in fact, the probability of relapse should concern only the category of offences. Of these, on a propositional level, one could use a second delimitation criterion that recurs many times in the criminal justice system. That is, one could circumscribe the object of the prognosis to crimes punishable by imprisonment above a certain threshold.

This could lead one to imagine the introduction of that greater burden of motivation on the judge to rebalance the decision taken.

Even in relation to the period of validity of the prognosis, more thought would be needed: it would seem to circumscribe it temporally: in fact, not an offence committed at any time in an indeterminate future, but a crime of a certain gravity within a time span not exceeding a certain threshold.

Moreover, more the prognosis is indeterminate in time, the less reliable are the criteria of rationality governing the judgement of the offender's future conduct.

In fact, this reflection is compounded by the fact that most empirical research on recidivism is carried out within a fixed period.

These is data from which one can derive an element, more or less reliable, on the a priori probability that a certain offender will reoffend. Indeed, the identification of a time limit for the formulation of the prognostic assessment seems necessary in order to be able to use the data compiled from existing empirical research on recidivism. The need to delimit this analysis precisely in time is linked to the risk and necessity of not losing an unspecified future and could also affect the controllability of the decision.

The most difficult point to unravel was also that of extracting oneself with difficulty from the idea that this type of reasoning can apply to all types of offences and that it leaves out instead those cases that are implemented as the result of a choice, of an impromptu reaction. However, an attempt has been made to sketch an overall picture and, above all, to try to answer the main question: 'is it possible/desirable to place technology tools, and in particular artificial intelligence, alongside the human being when he/she is faced with a decision that has prognostic characteristics'?

In an attempt to answer this question, an attempt was made to identify those tools that are and are currently used in this type of assessment.

Indeed, the analysis carried out shows how an entirely different world emerges in the transatlantic landscape: systems of justice that see such instruments applied daily at the moment in most pre-trial settings. This element, indeed, shows how the need and demand for greater precision and probably efficiency and celerity of the system has called these instruments into

play, without, however, envisaging ex ante a system and regulation capable of accommodating them without violating fundamental guarantees.

And it is precisely the perspective provided by overseas scholars and early effects that makes it, in some ways, easier to approach this issue in a critical manner since it allows one to have a sort of 'eye on the future' and assess, as of now, what the prospects might actually be. The continental European context, however, still looks at this new technological justice with curiosity and fear.

In the final part of the paper, an attempt has also been made to look at what represent, so to speak, real 'stumbling blocks' against which the possibility of introducing such instruments runs. In fact, in the development of Chapters IV and V, the most problematic and unresolved points detected in relation to the vague legislative formulations that require the judge to make a prediction emerge without doubt.

On closer inspection, the present study has attempted to look at the most recent and emerging literature on a topic that is still fairly new, attempting to assess the boundaries and the possibility of overcoming them by imagining a sort of 'compound assessment'.

Clearly, one cannot fail to say that such an analysis has been conducted and is still in an embryonic state that needs to be further investigated and developed.

What one would like to envisage is the possibility of placing side by side with the judging body, after having ascertained the guilt of an individual, a tool that takes into account certain factors (previously selected) and can issue a score that the judge can use in their decision, and then subsequently motivate the choice of the best sanctioning treatment.

Clearly, major risks that emerge in the latter part of this paper relate to the greater difficulty of setting boundaries in the use of such an instrument, of being able to control the effects and margins of error of such an instrument, and of ensuring that the adjudicating body would be able to provide and motivate.

In conclusion, it is therefore considered that this topic is so complex because it leads to questions not only about the nature of the concepts of deciding and ruling, but also because it leads to certain reflections that are difficult to unravel.

For these reasons, this analysis cannot stop at this necessary first step in approaching the subject; instead, it deserves and needs future in-depth studies that nevertheless call into play other branches, such as, for example, legal informatics, science, technology, and the philosophy of law.

A.I. can probably, as a result of the analysis carried out, support the judge and the judge for certain circumscribed aspects and procedural moments; however, the aspiration to

efficiency 'cannot entail a passive and total delegation to information technology of the exercise of judicial activity'⁸⁵⁵. In fact, the quality of justice is inextricably linked with the sensitivity, experience, and ability of the judge as an individual to grasp the small circumstances that make each decision unique and not manageable in a standardised or statistical manner'⁸⁵⁶.

Lastly, one realises, from the analysis made in the last chapter, how there is no effective uniform regulation on A.I.

In fact, in a future perspective that goes beyond and looks beyond this research that is still in its embryonic phase, it is deemed all the more necessary to set regulatory and normative limits in such a field in order to achieve a fruitful and guaranteed interaction between these two subjects that find themselves interacting for the first time; In this future perspective, a special form of regulation is deemed all the more necessary in order to be able to maintain respect for rights, in order to be able to assert new emerging rights, especially in the judicial phase; in fact, man relies on ethical choices and the machine on coherent and efficient choices that, however, do not always ensure fairness. This is because the machine is not capable of grasping the complex nuances of each situation and is at the same time unable to carry out a balancing of judgements and a gradation of the different interests at stake.

In conclusion, it is believed that the optimistic starting prospect of being able to 'yearn' or to strive for a 'just sentence' are perspectives that have been disregarded (or even that of an 'exact' prognostic assessment of criminal dangerousness that at the same time seems to constitute not only the wish of those who find themselves in criminal proceedings in the capacity of defendant, but the ambitious goal of human beings who ceaselessly strive to overcome boundaries that are always being moved further. What one could undoubtedly hope for, in an attempt to make the best possible use of such tools, is the prospect of a 'constitutionally oriented' use that does not sacrifice the fundamental guarantees in the face of information progress but, on the contrary, helps in the enhancement of certain decisions referred to the judging body and which could provide support in the '*dosimetry of punishment*'.

⁸⁵⁵As historian Yuval N. Harari writes in his topical essay "21 Lessons for the 21st Century" "Human beings have always been far better at inventing tools than using them wisely".

⁸⁵⁶D. POLIDORO, *Tecnologie informatiche e procedimento penale: la giustizia penale "messa alla prova" dall'intelligenza artificiale* in *Archivio Penale*, 2020, No. 3, 22

Bibliography

Books

- ALEO S. – DI NUOVO S., *Responsabilità penale e complessità. Il diritto penale di fronte alle altre scienze sociali. Colpevolezza, imputabilità, pericolosità sociale*, Milan, 2011;
- ALPA G. (ed), *Diritto e Intelligenza artificiale*, Pisa, 2020;
- ALPAYDIN E., *Introduction to Machine Learning*, Cambridge, 2010;
- ASHWORTH A., *Sentencing and Criminal Justice*, Cambridge, 2015;
- AVANZINI G., *Decisioni amministrative e algoritmi informatici*, Naples, 2019;
- BELLOTTI S. - MARIOTTI S. - MELONI A. - RUSSO R., *La certezza della pena. Nuove frontiere nel giudizio di pericolosità sociale*, Milan, 2020;
- BIANCHI A. – GULOTTA G. – SARTOR G. (eds), *Manuale di neuroscienze forensi*, Milan, 2009;
- BOVIO G., *Saggio critico del diritto penale*, Naples, 1883;
- BRICOLA F., *La discrezionalità nel diritto penale*, Milano, 1965;
- CADOPPI A., *Il valore del precedente nel diritto penale*, Turin, 1999;
- CALZOLAIO E., *La decisione nel prisma dell'intelligenza artificiale*, Milan, 2020;
- CAMPBELL J. C, *Assessing Dangerousness: Violence by sexual offenders, batterers, and child abusers*, New York, 1995;
- CANEPA M. – MERLO S., *Manuale di diritto penitenziario*, Milan, 1991;
- CANZIO G. – LUPARIA L. (eds), *Prova scientifica e processo penale*, Padua, 2018;
- CARLEO A. (ed), *Calcolabilità giuridica*, Bononia, 2017;
- CARRARA F., *Programma del Corso di diritto criminale*, Lucca, 1867;
- CASELLA A., *Le conseguenze sanzionatorie del reato*, Turin, 2011;
- CATERINI M. - ROMANO S. (eds), *Il sistema penale ai confini delle hard sciences. Percorsi epistemologici tra neuro-scienze e intelligenza artificiale*, Pisa, 2021;
- CERINI D. - PISANI A.T., *Smart mobility, smart cars e intelligenza artificiale: responsabilità e prospettive*, Turin, 2019;

CERQUETTI G., *L'imputabilità nella sistematica del diritto penale*, Perugia, 1970;

CHARPENTIER J., *Justice Machines. Racconto di fantascienza giudiziaria*, Macerata, 2015;

CONFORTI B. – RAIMONDI G., *Commentario alla convenzione europea dei diritti dell'uomo e delle libertà fondamentali*, Padua, 2002;

CONTISSA G., *Information technology*, Turin, 2017;

CORDERO F., *Codice di procedura penale commentato*, Turin, 1990;

ID., *Guida alla procedura penale*, Turin, 1986;

COSTANTINO F., *Autonomia dell'amministrazione e innovazione digitale*, Naples, 2012;

CUPELLI C., *La legalità delegata. Crisi e attualità della riserva di legge nel diritto penale*, Naples, 2012;

D'ACQUISTO G., *Intelligenza artificiale. Elementi*, Turin, 2021;

DE LEONARDIS F., *Big Data, decisioni amministrative e "povertà" di risorse della Pubblica amministrazione*, in E. Calzolaio (ed), *La decisione nel prisma dell'intelligenza artificiale*, Milan, 2020;

DELL'OSSO G., *Capacità a delinquere e pericolosità sociale*, Milan, 1985;

DOLCINI E., *La commisurazione della pena. La pena detentiva*, Padua, 1968;

DOMINGOS P., *L'algoritmo definitivo. La macchina che impara da sola e il futuro del nostro mondo*, Bollati Boringhieri, Turin, 2016;

DOMINIONI O., *La prova penale scientifica*, Milan, 2005;

DOUGLAS M. - WILDAVSKY A., *Risk and Culture: An essay on the selection of technical and environmental dangers*, Berkeley, 1983;

DUNI G., *L'amministrazione digitale. Il diritto amministrativo nella evoluzione telematica*, Rome, 1992;

DURANTE M., *Potere computazionale. L'impatto delle ICT su diritto, società e sapere*, Milan, 2019;

FANCHIOTTI V., *La giustizia penale statunitense. Procedure v. Antiprocedure*, Turin, 2022;

FARO S.-FROSINI T. E. - PERUGINELLI G., *Dati e algoritmi. Diritto e diritti nella società digitale*, Bononia, 2020;

FERGUSON G. A., *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, New York, 2017;

FERRUA P., *La prova nel processo penale. Struttura e procedimento*, Vol. I, Turin, 2015;

ID., *Il giusto processo* (3rd ed.), Bononia, 2012;

FIANDACA G., *Il diritto penale tra legge e giudice*, Padua, 2002;

FORNARI U., *Al di là di ogni ragionevole dubbio. Ovvero sulla cosiddetta prova scientifica nelle discipline psicoforensi*. Turin, 2012;

FRAGOLA S. P., *Le misure di prevenzione*, Padua, 1992;

FROSALI R. A., *Sistema penale italiano*, Vol. I, Turin, 1958;

GARAPON A. – LASSÈGUE J., *Justice digital. Revolution graphique et ropture anthropologique*, Paris, 2018;

GAROFALO E., *Criminologia. Studio sul delitto, sulle sue cause e sui mezzi di repressione*, Turin, 1861;

GIALUZ M., *Il carcere e la promessa tradita della costituente*, Bononia, 2012;

GIANNITI P., *La valutazione della prova penale*, Turin, 2005;

GRANDI C., *Riserva di legge e legalità penale europea*, Milan, 2020;

ID., *Neuroscienze e responsabilità penale*, Turin, 2016;

GROSSI P., *L'invenzione del diritto*, Bari-Roma, 2017;

GROSSO C. F. - NEPPI MODONA G. - VIOLANTE L., *Giustizia penale e poteri dello Stato*, Milan, 2002;

GROSSO C. F. – PADOVANI T. – PAGLIARO A., *Trattato di diritto penale*, Milan, 2009;

GUARNERI G., *Pericolosità sociale*, in *Noviss. Dig. It.*, vol. XII, Turin, 1965;

HAGE J. C., *Dialectical models in artificial intelligence and law*, in *Artificial Intelligence and law*, Berlin, 2000;

HARCOURT B., *Against prediction: Profiling, policing and punishing in actuarial age*, Chicago, 2007;

HASSEMER W., *Einführung in die Grundlagen des Strafrechts*, München 1981;

HEAVEN D. (ed), *Macchine che pensano. La nuova era dell'intelligenza artificiale*, Bari, 2018;

HELPER M. – RONCO M. (eds), *Diritto penale e autoresponsabilità. Tra paternalismo e protezione dei soggetti vulnerabili*, Turin, 2020;

IACOVIELLO F. M., *La Cassazione penale. Fatto, diritto e motivazione*, Milan, 2013;

JEAN A., *Nel paese degli algoritmi*, Neri Pozza editore, Vicenza, 2021;

KAHNEMAN D., *Pensieri lenti e veloci*, Milan, 2012;

KAHNEMAN D. – SIBONY O. – SUNSTEIN C., *Rumore. Un difetto del ragionamento umano*, Milan, 2021;

KAISER G., *Criminologia*, Milan, 1985;

KARANASIOU A., *Towards a Legal Definition of Machine Intelligence: The Argument for Artificial Personhood in the Age of Deep Learning*, in *Proceedings of ICAIL*, London, 2017;

KAPLAN J., *Intelligenza artificiale. Guida al futuro prossimo*, Rome, 2018;

KHUN T., *La Struttura delle rivoluzioni scientifica*, Turin, 1962;

LOMBROSO C., *Trattato antropologico sperimentale dell'uomo delinquente studiato in rapporto alla antropologia, alla medicina legale e alle discipline carcerarie*, Milan, 1876;

LOPEZ DE MANTARAS BADIA R. - MESEGUER COLZALEZ P., *Inteligencia artificial*, Madrid, 2017;

MANES V. – CAIANIELLO M., *Introduzione al diritto penale europeo*, Turin, 2020;

MANTOVANI F., *Manuale di Diritto penale*, Milan, 2020;

MARMO R., *Algoritmi per l'intelligenza artificiale. Progettazione dell'algoritmo - Dati e Machine Learning - Neural Network - Deep Learning*, Milan, 2020;

MARONGIU D., *L'attività amministrativa automatizzata*, Rimini, 2005;

MAYR E. - TOWAR, *A New Philosophy of Biology*, Cambridge, 1968;

MASSARO A., *La responsabilità colposa per omesso impedimento di un fatto illecito altrui*, Naples, 2013;

MASTRONARDI V. M., *Manuale per operatori criminologici e psicopatologi forensi*, Milan, 1996;

DE P., *Pericolosità sociale e legalità*, Milan, 2012;

MAYR E. -TOWAR, *A New Philosophy of Biology*, Cambridge, 1968;

MAZZACUVA F., *Le pene nascoste. Topografia delle sanzioni punitive e modulazione dello statuto garantistico*, Turin, 2017;

MESSINA S., *La discrezionalità nel diritto penale*, Rome, 1947;

MILITELLO V., *Prevenzione generale e commisurazione della pena*, Milan, 1982;

MILITELLO V. – SPENA A., *Mobilità, sicurezza e nuove frontiere tecnologiche*, in *Quaderni di Diritto penale comparato, internazionale ed europeo*, Turin, 2018;

MITCHELL T., *Machine Learning*, New York, 1997;

MONTAGNA M., *I confini dell'indagine personologica nel processo penale*, Rome, 2013;

NIEVA-FENOLL J., *Intelligenza artificiale e processo*, Turin, 2018;

O'NEIL C., *Weapons of Math Destruction. How Big Data Increases Inequality and threatens Democracy*, Penguin Books Ltd., New York, 2016;

PADOA SCHIOPPA A., *La giuria penale in Francia*, Milan, 1994;

PALAZZO F., *Corso di diritto penale. Parte generale*, Turin, 2018;

PALIERO C. E., *Il principio di effettività nel diritto penale*, Naples, 2011;

PASCERI G., *Intelligenza artificiale. Algoritmo e machine learning*, Milan, 2021;

PASQUALE F., *The black box society. The secret algorithms that control money and information*, London, 2015;

PEDRAZZI C., *Introduzione al diritto penale*, Milan, 603.

PELISSERO M., *Pericolosità sociale e doppio binario*, Turin, 2008;

PERRONE D., *La prognosi postuma tra distorsioni cognitive e software predittivi. Limiti e possibilità del ricorso alla “giustizia digitale integrata” in sede di accertamento della colpa*, Turin, 2022;

PETRINI D., *La prevenzione inutile*, Naples, 1996;

PORZIO L., *Sistemi punitivi e ideologie*, Naples, 1965;

QUINTARELLI S., *Intelligenza Artificiale: cos'è davvero, come funziona, che effetti avrà*, Turin, 2020;

RAMAJOLI S., *La prova nel processo penale*, Padua, 1995;

RECCHIA N., *Il principio di proporzionalità nel diritto penale. Scelte di criminalizzazione e ingerenza nei diritti fondamentali*, Turin, 2020;

RIONDATO S., *Robot: talune implicazioni di diritto penale*, in P. MORO-SARRA, *Tecnodiritto. Temi e problemi di informatica e robotica giuridica*, Milan, 2017;

ROMANO A. (ed), *L'azione amministrativa*, Turin, 2016;

RUFFOLO U. (ed), *Il diritto, i diritti, l'etica*, Milan, 2020;

RUMIATI R. – BONA C., *Dalla testimonianza alla sentenza. Il giudizio tra mente e cervello*, Bononia, 2018;

ROCCHI F., *La recidiva tra colpevolezza e pericolosità. Prospettive di indagine nel sistema penale integrato*, Naples, 2020;

ROMANO B., *Algoritmi al potere. Calcolo giudizio pensiero*, Turin, 2018;

ROMANO M. – GRASSO G., *Commentario sistematico del codice penale*, Vol. 2, Milan, 2012;

SADIN E., *Critica della ragione artificiale: una difesa dell'umanità*, Rome, 2019;

SANTANGELO A., *Precedente e prevedibilità. Profili di deontologia ermeneutica nell'era del diritto penale giurisprudenziale*, Turin, 2022;

SARTOR G., *Intelligenza Artificiale e diritto. Un'introduzione*, Milan, 1996;

SELVAGGI N. – CORTESI M. – LA ROSA E. – PARLATO I. – FLOR R. (eds), *Sistema penale e tutela delle vittime tra diritto e giustizia*, Milan, 2015;

SICLARI B., *Applicazione ed esecuzione delle misure di sicurezza personali*, Milan, 1977;

SCAGLIARINI S. (ed), *Smart roads and Driverless Cars: tra diritto, tecnologica, etica pubblica*, Turin, 2019;

SGUBBI F., *Il diritto penale totale*, Bononia, 2019;

SHAPIRO D. L. – NOE A. M., *Risk Assessment. Origins, Evolution, and Implications for Practice*, Berlin, 2015;

SINGH I. – SINNOT-ARMSTRONG W. P. - SAVULESCU J. (eds), *Bioprediction, Biomarkers, and Bad Behavior. Scientific, Legal, and Ethical Challenges*. Oxford, 2013;

SIPSER M., *Introduzione alla teoria della computazione*, in C. De Felice - L. Gargano – P. - D'arco (eds), Santarcangelo di Romagna, 2013;

SOURDIN T., *Judges, Technology and Artificial Intelligence. The artificial judge*, Cheltenham, 2021;

SPIRITO U., *Storia del diritto penale italiano da Cesare Beccaria ai giorni nostri*, Turin, 1932;

SUSSKIND R., *Online Courts and the Future of Justice*, Oxford, 2019;

TADROS V., 'Rethinking the Presumption of Innocence', in *Criminal Law and Philosophy*, 1, 2007;

TAGLIARINI F., (voce) *Pericolosità*, EDD, Milan, 1983;

TAMBURRINI G., *Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiale*, Rome, 2020;

TARONI M. – UBERTONE M., *Il diritto debole. Mutazione del diritto e nuove forme di normatività*, Turin, 2020;

TARUFFO M., *La semplice verità. Il giudice e la costruzione dei fatti*, Bari, 2009;

ID., *Sui confini. Scritti sulla giustizia civile*, Bononia, 2001;

TREZZA F., *Diritto e intelligenza artificiale, Etica, Privacy, Responsabilità, Decisione*, Pisa, 2020;

TURING A., *Mechanical Intelligence*, North Holland, 1992;

UBERTIS G., *Profili di epistemologia giudiziaria*, Milan, 2015;

ID., *Argomenti di procedura penale*, Vol. IV, Milan, 2016;

VASSALLI G., *Il potere discrezionale del giudice nella commisurazione della pena*, in *Primo corso di perfezionamento per uditori giudiziari*, II, Milan, 1958;

VENTUROLI M., *Modelli di individualizzazione della pena. L'esperienza italiana e francese nella cornice Europea*, Turin, 2020;

VESPIGNANI A., *L' algoritmo e l'oracolo: come la scienza predice il futuro e ci aiuta a cambiarlo*, Milan, 2019;

WARWICK K., *Intelligenza Artificiale. Le basi*, Palermo, 2015;

YEUG K. – LODGE M., *Algorithmic regulation*, Oxford, 2019;

ZAGREBELSKY V., *Manuale dei diritti fondamentali in Europa*, Bononia, 2019;

ZARA G., *Persistenza e recidivismo criminale: il risk assessment in psicologia criminologia*, Turin, 2010;

ZARA G. – FARRINGTON D., *Criminal recidivism: explanation prediction and prevention*, London, 2016.

Chapters in Edited Books

ALLEGREZZA S., *Prova scientifica e dimensione europea*, in G. Canzio – L. Luparia (eds), *La prova scientifica*, Milan, 2018;

ALLEGREZZA S. – COVOLO V., *Conclusions*, in S. Allegrezza – V. Covolo (eds), *Effective defence rights in criminal proceedings: a europea and comparative study on judicial remedies*, Padua, 2018;

AMISANO M., *Profetica-mente: neuroscienze, intelligenza artificiale e previsione*, in F. Basile-M. Caterini-S. Romano (eds), *Il sistema penale ai confini delle hard sciences*, Pisa, 2021;

BARTOLI R., *Diritto penale e prova scientifica*, in (a cura di) a cura di G. Canzio e L. Lupária, *Prova scientifica e processo penale*, Milan, 2018;

BASILE F. – CATERINI M. - ROMANO S. (a cura di), *Il sistema penale ai confini delle hard sciences. Percorsi epistemologici tra neuro-scienze e intelligenza artificiale*, Pisa, 2021;

BODEN M. A., *Intelligenza artificiale*, in J. I-Khalili (ed), *Il futuro che verrà*, Bollati Boringhieri, 2018;

BOSCO F. (et oths), *Profiling Technologies and Fundamental Rights and Values: Regulatory Challenges and Perspectives from European Data Protection Authorities*, in S. Gutwirth (et oths) (eds), *Reforming European Data Protection Law*, Berlin, 2015;

BICHI R., *Intelligenza digitale, giurimetria, giustizia predittiva e algoritmo decisorio. Machina sapiens e il controllo sulla giurisdizione*, in U. Ruffolo (ed), *Intelligenza artificiale. Il diritto, i diritti, l'etica*, Milano, 2020;

BURRELL W.D., *Risk and Needs Assessment in Probation and Parole: The Persistent Gap Between Promise and Practice*, in F. Taxman (ed) *Handbook on Risk and Need Assessment: Theory and Practice*, New York, 2017;

CHIAVARIO M., *Art. 6*, in S. Bartole - B. Conforti - G. Raimondi (eds), *Commentario alla convenzione europea dei diritti dell'uomo e delle libertà fondamentali*, Padua, 2002;

COVELLI M. R., *Dall'informatizzazione della giustizia alla «decisione robotica»? Il giudice del merito*, in A. Carleo (ed), *Decisione robotica*, Bononia, 2019;

CAVALIERE A., *Può la "sicurezza" costituire un bene giuridico o una funzione del diritto penale?*, in *In dubio pro libertate*, in W. Hassmer – E. Kempf – S. Moccia (eds), *Festschrift für Klaus Volk zum 65. Geburtstag*, München, 2009;

FERRUA P., *Epistemologia scientifica ed epistemologia giudiziaria: Differenze, analogia, interrelazioni*, in L. De Cataldo Neuburger (ed), *La prova scientifica nel processo penale*, Padua, 2007;

FERRUA P., *Commento all'art. 27, comma 3*, in *Commentario alla Costituzione. Rapporti civili*, in G. Branca - A. Pizzorusso (eds), Bononia, 1991;

FLOR R., *Le nuove frontiere del contrasto alla criminalità: dalle investigazioni tecnologiche alla predictive policing al servizio della Urban Security*, in di T. Dalla Massara-M. Beghini (eds), *La città come bene comune*, Naples, 2019;

GILLESPIE T., *The relevance of Algorithms*, in T. Gillespie - P. Boczkowski - K. Foot (eds.), *Media Technologies*, Cambridge, 2014;

GRANDI C., *El papel de la neurociencia en el juicio de la imputabilidad: el debate teorico y las consecuencias practicas en la experiencia italiana*, in E. D. Crespo (ed), *Avances desde la neurociencia y la inteligencia artificial*, Barcelona, 2022;

ID., *Recensione a “Colpevolezza, conseguenze sanzionatorie e neuroscienze in rapporto al diritto penale*, in M. Di Florio (ed), Pisa, 2020;

HILDEBRANDT M., *Profiling and AML*, in K. Rannenber (et oths) (eds), *The Future of Identity in the Information Society. Challenges and Opportunities*, Berlin, 2009;

ITALIANO G. F., *Intelligenza artificiale: passato, presente, futuro*, in F. Pizzetti (ed), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Turin, 2018;

LANGBEIN J. H., *The English Criminal Trial Jury on the Eve of the French Revolution*, in A. Padoa Schioppa (eds), *The Trial Jury in England, France and Germany, 1700-1900*, Berlin, 1987;

LASAGNI G., *Difendersi dall’intelligenza artificiale o difendersi con l’intelligenza artificiale? Verso un cambio di paradigma*, in G. Di Paolo – L. Pressacco (eds), *Intelligenza artificiale e processo penale: indagini, prove e giudizio*, Naples, 2022;

LUCIANI M., *La decisione giudiziaria robotica*, in A. Carleo (ed), *Decisione robotica*, Bononia, 2019;

MAGGIO P., *La risocializzazione e la tutela della dignità dello straniero minorenne devono prevalere sulle finalità delle espulsioni*, in S. Greco – G. Tumminelli (eds), *Migrazioni in Sicilia*, Milan, 2020;

ID., *La portata delle garanzie difensive nel rinnovato ordinamento penitenziario*, in S. Lorusso (ed) *Il fragile mosaico delle garanzie difensive*, Turin, 2020;

MANES V. – MAZZACUVA F., *Le disposizioni penali*, in G. Finocchiaro (ed), *La protezione dei dati personali in Italia*, Bononia, 2019;

MANZINI V., *Trattato di diritto penale italiano*, P. Nuvolone – G. D. Pisapia (eds) vol. III, Turin, ed. 1981;

MAZZACUVA F., *La Convenzione europea dei diritti dell’uomo e i suoi riflessi sul sistema penale*, in A. Cadoppi – S. Canestrari – A. Manna – M. Papa (eds), *Trattato di diritto penale. Parte generale*, Turin, 2012;

ID., *L’evoluzione nazionale ed internazionale della confisca tra diritto penale “classico” e diritto penale “moderno”*, in A. Bargi – A. Cisterna (eds), *Giustizia patrimoniale penale*, Turin, 2011;

- MILETO P., *Le misure di prevenzione*, in G. Ambrosini - P. Miletto (eds), *Le sostanze stupefacenti. Le misure di prevenzione*, in *Giurisprudenza sistematica di diritto penale*, F. Bricola – V. Zagrebelsky (eds), Turin, 1989;
- MORIGNAT V., *L'I.A., dalle previsioni alle decisioni*, in *L'intelligenza artificiale tra etica e diritti. Prime riflessioni a seguito del libro bianco dell'Unione europea*, in A.F. Uricchio – G. Riccio – U. Ruffolo (eds), Bari, 2020;
- PAGALLO U., *Etica e diritto dell'Intelligenza Artificiale nella governance del digitale: il Middle-out Approach*, in *Intelligenza artificiale*, in U. Ruffolo (ed), *Intelligenza artificiale. Il diritto, i diritti, l'etica*, Milan, 2020;
- PAGALLO U. – DURANTE M., *The Philosophy of Law in an Information Society*, in L. Floridi (ed.), *The Routledge Handbook of Philosophy of Information*, New York, 2016;
- PALMIRANI M., *La mediazione digitale e nuove forme tecnologiche di discriminazione*, in AA.VV. (eds), *La mediazione interculturale*, Bononia, 2021;
- PAPA M., *Future crimes: intelligenza artificiale e rinnovamento del diritto penale*, in S. Dorigo (a cura), *Il ragionamento giuridico nell'era dell'intelligenza artificiale*, Pisa 2020;
- PASCULLI L., *Genetics, Robotics and Crime Prevention*, in D. Provolo – S. Riondato – F. Yenisey (eds), *Genetics, Robotics and Punishment*, Padua, 2014;
- PERCHINUNNO V., “*Le prove*”, in M. Pisani (eds), *Manuale di procedura penale*, Bononia, 2008;
- PUNZI A., *Judge in the Machine: e se fossero le macchine a restituirci l'umanità del giudicare?*, in Carleo A. (ed), *Decisione robotica*, Bononia, 2019;
- ROMANO G., *Diritto, robotica e teoria dei giochi: riflessioni su una sinergia*, in *Diritto e intelligenza artificiale*, in G. Alpa (ed), *Intelligenza artificiale, giustizia penale, controllo umano significativo*, Pisa, 2020;
- SARTOR G. – LAGIOIA F., *Le decisioni algoritmiche tra etica e diritto*, in (a cura di) U. Ruffolo, *Intelligenza artificiale. Il diritto, i diritti, l'etica*, Milan, 2020;
- SCARPELLI U., *Gli orizzonti della giustificazione*, in L. Gianformaggio - E. Lecaldano (eds), *Etica e Diritto. Le vie della giustificazione razionale*, Roma-Bari, 1986;
- TARELLO G., *L'interpretazione della legge*, in AA.VV., *Trattato di diritto civile e commerciale*, Milan, 1980;

THAMAN S., *Should criminal juries give reasoning for their verdicts? Th Spanish experience and the implications of the European Court of Human Rights decision in Taxquet v. Belgium*, in A. Petrova (ed), *Festschrift für August Nacke*, 2016;

TRIPODI A. F., *Abusi di mercato e trading algoritmico*, in AA.VV., *Il diritto nell'era digitale*, Milan, 2022;

UBERTIS G., *La prova scientifica e la nottola di Minerva*, in L. De Cataldo Neuburger (ed.), *La prova scientifica nel processo penale*, 83–91, Padua, 2007;

Articles

AARAUJO T. (et oths), *In AI we trust? Perceptions about automated decision-making by artificial intelligence*, in *AI & Society*, 35, 2020;

ALDARONDO E. – D. B. SUGARMAN, *Risk marker analysis of the cessation and persistence of wife assault*, in *Journal of consulting and clinical psychology*, 64(5), 1996;

ALLEN C. – VARNER G. – ZINSER J., *Prolegomena to any future artificial moral agent*, in *Journal of Experimental & Theoretical Artificial Intelligence*, 2000;

ALTAVILLA E., *Studi sul progetto del nuovo codice penale. Visione positivista della parte speciale di un nuovo codice criminale*, in “*Scuola Positiva*”, 1921;

ANDREWS D. A. – BONTA J., *Rehabilitating criminal justice policy and practice*, in *Psychology, Public policy and Law*, 16(1) 2010;

ARBOTTI M., *Recensione a Il diritto, la giustizia e lo spazio. Note a margine di A. Garapon, La despazializzazione della giustizia*, in *Sistema penale*, 2022;

ARDUINI S., *La “scatola nera” della decisione giudiziaria: tra giudizio umano e giudizio algoritmico*, in *Riv. Biolaw*, no. 2/2021;

AVERY J. J., *An Uneasy Dance with Data: Racial Bias in Criminal Law*, in *Southern California Law Review Postscript*, Vol. 93, No. 28, 2020;

BALKIN J.M., *The Path of Robotics Law*, in *California Law Review Circuit*, Vol. 6, 2015;

BARBARO C., *Usò dell'Intelligenza artificiale nei sistemi giudiziari: verso la definizione di principi etici condivisi a livello europeo?*, in *Quest. Giust.*, no. 4/2018;

- BAROCAS S. – SELBST A. D., *Big Data's Disparate Impact*, in *California Law Review*, Vol. 104, No. 3, 2016;
- BASILE F., *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Diritto penale e uomo*, 23 September 2019;
- BECK S., *Intelligent agents and criminal law - Negligence, diffusion of liability and electronic personhood*, in *Robotics and Autonomous Systems*, 2016;
- M. BELKIN – D. HSU – S. MA – S. MANDAL, *Reconciling modern machine learning practice and the classical bias-variance trade off*, in *PNAS*, Vol. 116, 32, 2019;
- BELLAVISTA G., *Il potere discrezionale nell'applicazione della pena*, 1939, in *Il Tommaso Natale*, 1975;
- BERK R. – HYATT J., *Machine Learning Forecasts of Risk to Inform Sentencing Decisions*, in *HeinOnline*, 2015;
- BERLINGÒ V., *Datafication e giuridicizzazione*, in *Riv. Tri., dir. Pubbl.*, 2017;
- BERTOLINO M., *Declinazioni attuali della pericolosità sociale: pene e misure di sicurezza a confronto*, in *Arch. Pen.*, 2014;
- ID., *Profili vecchi e nuovi della imputabilità e della sua crisi*, in *Riv. It. Dir. Proc. Pen.*, 1988;
- BETTIOL G., *Pena retributiva e poteri discrezionali del giudice*, in *Riv. it. dir. pen.*, 1941;
- BIELLI D., *Periti e consulenti tecnici nel nuovo processo penale*, in *Giustizia penale*, 1991;
- BONFANTI A., *Big data e polizia predittiva: riflessioni in tema di protezione del diritto alla privacy e dei dati personali*, in *Rivista di diritto dei media*, no. 3/2018;
- BONTA J. - LAW M. – HANSON K., *The prediction of criminal and violent recidivism among mentally disordered offenders: A meta analysis*, in *Psychological Bulletin*, 1998;
- BOTNICK C., *Evidence-Based Practice and Sentencing in State Courts: A Critique of the Missouri System*, in *HeinOnline*, 49 Wash. U. J. L. & Pol'y 159, 2015;
- BRANTINGHAM P. J., *The Logic of Data Bias and its Impact on Place-Based Predictive Policing*, in *HeinOnline*, 2018;
- BRASIELLO U., (voce) *Diritto penale (diritto romano)*, in *Noviss. dig. it.*, V, Turin, 1964;

BRENNAN-MARQUEZ K., *Big Data Poling and the Redistribution od Anxiety*, in *HeinOnline*, 2018;

BRUSCO C., *Scienza e processo penale: brevi appunti sulla valutazione della prova scientifica*, in *Riv. It. Med. Leg.*, No. 1/2012;

BURCHARD C., *L'intelligenza artificiale come fine del diritto penale? Sulla trasformazione algoritmica della società*, in *Riv. it. dir. proc. pen.*, 2019;

BURRELL J., *How the machine 'thinks': Understanding opacity in machine learning algorithms*, in *Big Data & Society*, June 2016;

CABIALE, A., *L'accertamento giudiziale della pericolosità tra presente e futuro*, in *Arch. Pen.*, No. 2, 2022;

CAIANIELLO M., *Dangerous Liasons. Potentialities and Risks Deriving from the Interaction between Artificial Intelligence and Preventive Justice*, in *European Journal of Crime, Criminal Law and Criminal Justice*, 2021, No. 1;

CALABRESI G.-AL MUREDEN E., *Driverless car e responsabilità civile*, in *Rivista di Diritto Bancario*, No. 3, 2020;

CAMPBELL J. C., *Nursing assessment of risk of homicide for battered women*. *Advances in Nursing Science*, 1986;

CANZIO G., *Il dubbio e la legge*, in *Arch. Pen.*, 20 July 2018;

CAPERS I. B., *Techno-Policing*, in *HeinOnline*, in *Ohio St. J. Crim. L.* 495, 2018;

CAPPELLINI A., *Profili penalistici delle self-driving cars*, in *Riv. Trim – Dir. Pen. Cont.*, No. 2, 2019;

CAPRIOLI F., *La scienza "cattiva maestra": le insidie della prova scientifica nel processo penale*, in *Cass. Pen.*, 2008;

CARCANO A., *Automatismi: tra ragionevolezza e individualizzazione della pena*, in *Forum di Quaderni costituzionali – Rassegna*, 4, 2021;

CARLONI E., *I Principi della legalità algoritmica. Le decisioni automatizzate dj fronte al giudice amministrativo*, in *Dir. Amm.*, 2020;

CARNELUTTI F., *Matematica e diritto*, in *Riv. Dir. Proc.*, 1951;

CARROZZA M.C. et others., *AI: profili tecnologici. Automazione e Autonomia: dalla definizione alle possibili applicazioni dell'Intelligenza Artificiale*, in *BioLaw Journal – Rivista di BioDiritto*, no. 3/2019;

CARULLO G., *Big data e pubblica amministrazione nell'era delle banche interconnesse*, in *Concorrenza e mercato*, 2016;

CASTELLI C. – PIANA D., *Giustizia predittiva. La qualità della giustizia in due tempi*, in *Quest. Giust.*, No. 4, 2018;

CASONATO C., *Intelligenza artificiale e diritto costituzionale: prime considerazioni*, in *Diritto pubblico comparato e europeo*, May 2019;

CATALANO E. M., *Logica della prova, statistical evidence e applicazione della teoria delle probabilità nel processo penale*, in *Riv. Trim. Dir. Pen. Cont.*, No. 4, 2013;

CATERINI M., *Il giudice penale robot*, in *La legislazione penale*, 2020;

CATH C. – WACHTER S. – MITTELSTADT B. – TADDEO M. – FLORIDI L., *Artificial Intelligence and the “Good Society”: the US, EU, and UK approach*, in *Science and Engineering Ethics*, 2017;

CAVALLO PERIN R., *L'amministrazione pubblica con i big data: da Torino un dibattito sull'intelligenza artificiale*, in *Quaderni di dipartimento dell'Università di Torino*, 2021;

CESARI C., *Editoriale: L'impatto delle nuove tecnologie sulla giustizia penale – un orizzonte denso di incognite*, in *Revista brasileira de direito processual penal*, Porto Alegre, Vol. 4, No. 3, 2019;

CEVOLANI G. – CRUPI V., *Come ragionano i giudici: razionalità, euristiche e illusioni cognitive*, in *Criminalia*, 2017;

CHIAO V., *Predicting Proportionality: The Case for Algorithmic Sentencing*, in *Criminal Justice Ethics*, Vol. 37, No. 3, 2018;

CLAUSSÉN M. -KARLSSON, *Artificial Intelligence and the External Element of the Crime An Analysis of the Liability Problem. An Analysis of the Liability Problem*, in *Juridicum*, 2017;

COLLICA M. T., *Il giudizio di imputabilità tra complessità fenomenica ed esigenze di rigore scientifico*, in *Riv. It. Dir. Proc. Pen.*, 2008,

COMELLA C., *Origine dei “Big data”*, in *Gnosis*, 2017;

CONSULICH F., *Il nastro di Möbius. Intelligenza artificiale e imputazione penale nelle nuove forme di abuso del mercato*, in *Banca Borsa Titoli di Credito*, Vol. 2, 2018;

CONTISSA. G – LASAGNI G. – SARTOR G., *Quando a decidere in materia penale sono (anche) algoritmi e IA: alla ricerca di un rimedio effettivo*, in *Riv. trim. diritto di internet*, No. 4, 2019;

CORDOVA A., *Le riforme della legislazione penale e il loro momento storico*, in *Rivista Penale*, 1921;

COSTANTINO F., *Lampi, nuove frontiere delle decisioni amministrative*, in *Dir. Amm.*, 2017;

COSTANZI C., *La matematica del processo: oltre le colonne d'Ercole della giustizia penale*, in *Questione Giustizia*, no. 4/2018;

CRISCI S., *Intelligenza artificiale ed etica dell'algoritmo*, in *Foro amm.*, 2018;

CRISTI S., *Evoluzione tecnologica e trasparenza nei procedimenti "algoritmici"*, in *Diritto di internet*, 2019;

CUPELLI C., *Sindacato costituzionale e discrezionalità legislativa*, in *Archivio DPC*, 27 March, 2019;

ID., *Hobbes europeista? Diritto penale europeo, auctoritas e controlimiti*, in *Discrimen*, 2 September 2018;

ID., *Il problema della legalità penale. Segnali in controtendenza sulla crisi della riserva di legge*, in *Giur. Cost.*, 2015;

D'AGOSTINO L., *Gli algoritmi predittivi per la commisurazione della pena*, in *Dir. Pen. Cont. – Riv. Trim.*, No. 2/2019;

DELGADO M., *Automazione, intelligenza artificiale e pubblica amministrazione: vecchie categorie concettuali per nuovi problemi?*, in *Istituzioni del Federalismo*, No. 3, 2019;

DELL'ANDRO R., *Il dibattito delle scuole penalistiche*, in *Arch. Pen.*, 1958;

DELOGU T., *Potere discrezionale del giudice e certezza del diritto*, in *Riv. it. dir. proc. pen.*, 1976;

DE MIGUEL BERIAIN I., *Does the use of risk assessments in sentences respect the right to due process? A critical analysis of the Wisconsin v. Loomis ruling*, in *Law, Probability and Risk*, 2018;

DESKUS C., *Fifth Amendment Limitations on Criminal Algorithmic Decision-Making*, in *HeinOnline*, 21 *N.Y.U. J. Legis. & Pub. Pol'y* 237, 2018;

DI GIOVINE O., *Il judge-bot e le sequenze giuridiche in materia penale (intelligenza artificiale e stabilizzazione giurisprudenziale)*, in *Cass. pen.*, 2020;

DI PRISCO A., *Elementi di criticità sulla perizia psicologica nel processo penale*, in *Ius in Itinere*, 2018;

ID., *I presupposti della responsabilità penale tra diritto e scienze*, in *Arch. Pen.* 22 June 2018;

DONATI F., *Intelligenza artificiale e giustizia in Rivista Associazione italiana dei Costituzionalisti*, in *Rivista AIC*, 1/2020;

DOYLE C. – C. BAINS – B. HOPKINS, *Principles of pretrial release: reforming bail without repeating its harms*, in *The Journal of Criminal Law and Criminology*, Vol 108, No. 4, 1973;

DUAN Y. - EDWARDS J. S. – DWIVEDI Y. K., *Artificial intelligence for decision making in the era of Big Data. Evolution, challenges and research agenda*, in *International Journal of Information Management*, 48, 2019;

DOUGLAS K. S. – SKEME J. L., *Violence risk assessment: getting specific about being dynamic*, in *Psychology Public Policy and Law*, September 2005;

DUNI G., voce *Amministrazione digitale*, in *Enc. Dir., Annali*, vol. I, 2007, Rome;

DUTTON D. G. – KROPP P. R., *A Review of domestic violence risk instruments*, in *Sage Journals*, Vol. 1, Issue 2, 2000;

ROEHL J. – GUERTIN K., *The current use of risk assessments in sentencing offenders*, in *The Justice systems journal*, Vol. 21, No. 2, 2000;

ETIENNE M., *Legal and Practical Implications of Evidence-Based Sentencing by Judges*, 1 *Chapman J. Crim. Just.* 43, 2009;

EUBANKS V., *Digitizing the carceral State, Automating inequality: how high-tech tools profile, police, and punish the poor*, in *Book Review*, 2016;

EUSEBI L., *La pena tra necessità di strategie preventive e nuovi modelli*, in *Riv. It. Dir. Proc. Pen.*, 2021;

ID., *Tra crisi dell'esecuzione penale e prospettive di riforma del sistema sanzionatorio: il ruolo del servizio sociale*, in *Riv. it. dir. pr. pen.*, 1993;

FALCONE M., *Big data e pubbliche amministrazioni: nuove prospettive per la funzione conoscitiva pubblica*, in *Riv. Trim. Dir. Pubbl.*, 2017;

ID., *Le potenzialità conoscitive dei dati amministrativi nell'era della "rivoluzione dei dati": Il caso delle politiche di eradicazione dell'epatite C*, in *Istituzioni del federalismo*, 2017;

FERGUSON G. A., *Illuminating Black Data Policing*, in *HeinOnline*, 15 Ohio St. J. Crim. L. 503, 2018;

FERRARA E. – VAROL O. – DAVIS C. – MENCZER F. – FLAMMINI A., *The Rise of Social Bots*, in *Review articles*, vol. 59, no. 7, July 2018;

FERRI E., *Relazione sul progetto preliminare di Codice penale italiano*, in *Scuola Positiva*, 1929;

FLORIDI L. – SANDRES J.W., *On the Morality of Artificial Agents*, in *Information Ethics Group*, in *Minds and Machines*, 2004;

FRONZA E. – CARUSO C., *Ti faresti giudicare da un algoritmo? Intervista ad Antoine Garapon*, in *Quest. Giust.*, 4, 2018;

GALETTA D.U. – CORVALAN J. G., *Intelligenza artificiale per una pubblica amministrazione 4.0? Potenzialità, rischi e sfide della rivoluzione tecnologica in atto*, in *Federalismi.it*, no. 3/2019;

GABORIAU S., *Libertà e umanità del giudice: due valori fondamentali della giustizia. La giustizia digitale può garantire nel tempo la fedeltà a questi valori?*, in *Quest. Giust.*, No. 4, 2018;

GABRIELLI E. – RUFFOLO U., *Dottrina e attualità giuridiche. Intelligenza artificiale e diritto*, in *Giur. It.*, 2019;

GALETTA D.U. – CORVALAN J.G., *Intelligenza Artificiale per una Pubblica Amministrazione 4.0? Potenzialità rischi e sfide della rivoluzione tecnologica in atto*, in *Federalismi.it*, 6 February, No. 1, 2019.

GALGANI B., *Considerazioni sui "precedenti" dell'imputato e del giudice al cospetto dell'IA nel processo penale*, in *Sist. pen.*, No. 4, 2020;

GALLO E., *L'evoluzione del pensiero della Corte costituzionale in tema di funzione della pena*, in *Giur. cost.*, 1994;

GARRETT B. L.– MONAHAN J., *Judging Risk*, in *California Law Review*, Duke, 2020;

GASTALDO F. T., *Il giudice-robot: l'intelligenza artificiale nei sistemi giudiziari tra aspettative ed equivoci*, in *Iusinitinere*, 22 March 2021.

GENDRAU P. - LITTLE T. – GOGGIN C., *A meta-analysis of the predictors of adult offender recidivism: What works!*, in *Criminology*, 1996;

GERARDS J., *The fundamental rights challenges of algorithms* in *Sage Journal NQHR*, 2019;

GIALUZ M., *Quando la giustizia penale incontra il diritto penale: luce e ombre dei risk assessment tools tra Stati Uniti e Europa*, in *Arch. Pen.*, 29 May 2019;

ID., *Gli automatismi cautelari tra legalità costituzionale e garanzie convenzionali*, in *Proc. Pen. E giust.*, 2013;

GIANNINI M. S., *Rapporto sui principali problemi dell'Amministrazione sullo Stato*, in *Riv. Trim. dir. Pubbl.*, 1982;

GIUFFRIDA I. – LEDERER F. – VERMEYS N., *A legal perspective on the trials and tribulations of A.I.: how artificial intelligence, the Internet of Things, Smart Contracts, and other technologies will affect the Law*, in *Case Western Reserve Law Review*, Vol. 68, Issue 3, 2019;

GIUNTA F., *Verso una nuova pericolosità sociale*, in *Cultura e diritti*, 2012;

GLAZEBROOK, J. S., *Risky business: Predicting recidivism*, in *Psychiatry, Psychology and Law*, Vol. 17, Issue 1, 2010;

GORLA G., *Precedente giudiziario*, in *Enc. Giur. Treccani*, vol. XXXVI, 1991;

GRANDI C., *Criminal law and neuroscience: theory and practice in the Italian perspective*, in *International Journal of Criminology and sociology*, Vol. 11, No. 1, 2022;

ID., *Neuroscienze e capacità di intendere e volere: un percorso giurisprudenziale*, in *Dir. Pen. Proc.*, No. 1, 2020;

GRANDI C., *Diritto penale e neuroscienze. Punti fermi (se mai ve ne siano) e questioni aperte*, in *DPU*, Vol. 2019, No. 4, 2019;

ID., *Sui rapporti tra diritto penale e neuroscienze*, in *Riv. It. Dir. Proc. Pen.*, Vol. 57, No. 3, 2014;

- GRANN M. - H. BELFRAGE - A. TENGSTROM, *Actuarial assesment of risk for violence: Predictive validity of the VRAG and the historical partof the HCR-20*, in *Criminal Justice and Behavior*, 2000;
- GRIGNETTI F., *Il giudice del futuro sarà l'algoritmo. L'intelligenza artificiale in aula*, in *Ristretti Orizzonti*, 18 October 2018;
- GROVE W. M. – ZALD D. H. - LEBOW B. S. – SNITZ B. E. – NELSON C., *Clinical versus mechanical prediction: A meta-analysis*. *Psychological Assessment*, 12, 19–30, 2000;
- GUARRIELLO V., *L'intelligenza artificiale tra profili giuridici ed alcune delle più attuali applicazioni al servizio della società*, in *ARSG*, 19 November 2021;
- GULLO A., *Nuove frontiere tecnologiche e sistema penale: alcune note introduttive*, in *Riv. Trim. - Dir. Pen. Cont.*, 2, 2019.;
- GULOTTA G., *A proposito di scienza e processo, commento a sentenza Sez. III, sent. 18 febbraio 2020*, in *Sistema penale*, 2 aprile 2021;
- HALLEVY G., *Dangerous Robots – Artificial Intelligence vs. Human Intelligence*, in *SSRN*, 2018;
- ID., *The Criminal Liability of Artificial Intelligence Entities - From Science Fiction to Legal Social Control*, in *HeinOnline*, 4 Akron Intell. Prop. J. 171, 2010;
- HAMILTON M., *Risk-Needs Assessment: Constitutional and Ethical Challenges*, in *HeinOnline*, 52 Am. Crim. L. Rev. 231, 2015;
- HANSON R. K. – BUSSIÈRE M. T. , *Predicting relapse: A meta-analysis of sexual offender recidivism studies*, in *Journal of Clinical and Consulting Psychology*, No. 66, 1998;
- HELPER M., *The principle of self-responsibility as a liability-limiting criterion proceedings*, in *International Snow Science Workshop*, Innsbruck, 2018;
- HENDERSON S. E., *A Few Criminal Justice Big Data Rules*, in *HeinOnline*, 15 Ohio St. J. Crim. L. 527, 2018;
- HILTON N. Z. – HARRIS G. T. – RICE M. E. , *Predictive violence by serious wife assoluters*, in *Sage Journal*, Vol. 16, Issue 5, 2001.
- HORSEFIELD A., *Risk assessment: Who needs it?*, in *Probation Journal*, 2003;

HOLSINGER A.M. (et oths), *A Rejoinder to Dressel and Farid: New Study Finds Computer Algorithm is More Accurate than Humans at Predicting Arrest and as Good as a Group of 20 Lay Experts*, Vo. 82, No. 2, September 2018;

HU M., *Algorithmic Jim Crow*, in *HeinOnline*, 86 Fordham L. Rev. 633, 2017;

HUQ A. Z., *Racial equità in algorithmic criminal justice*, in *Duke Law Journal*, Vol. 68, no. 6, March 2019;

INTRIERI C., *Neuroscienze e diritto: una Nuova teoria giuridica sulla mente*, in *Sistemi intelligenti*, XXII, No. 2, August 2010;

ISAAC W. S., *Hope, Hype, and Fear: The Promise and Potential Pitfalls of Artificial Intelligence in Criminal Justice*, in *HeinOnline*, 15 Ohio St. J. Crim. L. 543, 2018;

JOH E., *Automated Policing*, in *HeinOnline*, 15 Ohio St. J. Crim. L. 559, 2018;

KATYAL S. K., *Private Accountability in the Age of Artificial Intelligence*, in *HeinOnline*, 66 UCLA L. Rev. 54, 2019;

KATS Y., *Manufacturing an Artificial Intelligence Revolution*, November 2017;

KEHL D.-GUO P.-KESSLER S., *Algorithms in the Criminal Justice System: Assessing the use of Risk Assessments in Sentencing*”, *Responsive Communities Initiative, Berkman Klein Center for Internet and Society* (Harvard Law School, <https://dash.harvard.edu>), 2017;

KING T. C. - AGGARWALL N. - TADDEO M. –FLORIDI L., *Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions*, in *SSRN*, 2018;

KLINGELE C., *The Promises and Perils of Evidence-Based Corrections*, in *91 Notre Dame L. Rev.*, 2016;

KOSTORIS R., *Predizione decisoria, diversion processuale e archiviazione*, in *Sistema penale*, 23 July 2021;

KROLL J. A. –HUEY J. –BOROCAS S. –FELTEN E. W. –REIDENBERG J. R. –ROBINSON D. G. –YU H., *Accountable algorithms*, in *University of Pennsylvania Law Review*, vol. 165:633, 2017;

KROPP P. R., *Intimate partner violence risk assessment and management. Violence and Victims*, In *PubMed*, 2008;

- KUTATELADZE B., *Cumulative Disadvantage: Examining Racial and Ethnic Disparity in prosecution and sentencing*, in *Criminology: An Interdisciplinary Journal*, Vol. 52, No.3, 2014;
- LARIZZA S., *La commisurazione della pena. Rassegna di dottrina e giurisprudenza*, in *Riv. it. dir. proc. pen.*, 1982;
- LIGHTBOURNE J., *Damned Lies & Criminal Sentencing Using Evidence-Based Tools*, in *HeinOnline*, 15 Duke L. & Tech. Rev. 327, 2017;
- LIMITI C., *Intelligenza Artificiale: implicazioni etiche in materia di privacy e diritto penale*, in *Ius in itinere*, 9 February 2021;
- LOGG J. M. - MINSON J. A. – MOORE D. A., *Algorithm appreciation: People prefer algorithmic to human judgment*, in *Organizational Behavior and Human Decision Processes*, 2019;
- MADDALENA M. L., *La digitalizzazione della vita dell'amministrazione e del processo*, in *Foro amm.*, 2016;
- MAGGIO P., *La "canalizzazione" dell'impugnazione cautelare per l'offeso postulante nei reati violenti*, in *Il foro italiano*, 2022;
- ID., *Rapporti familiari e tutela processuale penale*, in *Proc. Pen. E giust.*, 2019;
- ID., *La Corte europea dei diritti dell'uomo promuove una versione debole del diritto di accesso al difensore?*, in *Cass. Pen.*, 2019;
- MAGI R., *Per uno statuto unitario dell'apprezzamento della pericolosità sociale. Le misure di prevenzione a metà del guado?*, in *DPC*, 3/2017;
- MAGRO B., *Biorobotics, robotics and criminal law: some hints and reflections*, in *Percorsi Costituzionali*, 2016;
- MALCOLM M. FEELEY - JONATHAN SIMON, *The New Penology: Notes on the Emerging Strategy of Corrections and its Implications*, in *Criminology*, 1st January, Vol. 30, No. 4, 1992;
- MALDONATO L., *Risk assessment tools e riforma del sistema sanzionatorio*, in *Discrimen*, 18 October, 2022;
- ID., *Algoritmi predittivi e discrezionalità del giudice: una nuova sfida per la giustizia penale*, in *Dir. Pen. Cont. – Riv. Trim.*, 2/2019;

- MALINVERNI A., *Capacità a delinquere*, in *Enc. Dir.*, Vol. VI, Milan, 1960;
- MANES, V., *Il “costo delle garanzie” nel “modello penale liberale”: tra regressioni culturali e lacerazioni congiunturali*, in *L’Indice penale*, No. 1, 2021;
- ID., *Diritto penale no-limits. garanzie e diritti fondamentali come presidio per la giurisdizione*, in *Quest. giust.*, 2019;
- ID., *Proporzione senza geometria*, in *Giur. cost.*, 2016;
- ID., *Il ruolo “poliedrico” del giudice penale, tra spinte di esegesi adeguatrice e vincoli di sistema*, in *Cass, pen.*, No. 5, 2014;
- ID., *L’oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in *Discrimen*, th May 2020;
- ID., *Il ruolo “poliedrico” del giudice penale, tra spinte di esegesi adeguatrice e vincoli di sistema*, in *Cass. pen.*, 2014;
- MANES V. – MAZZACUVA F., *GDPR e nuove disposizioni penali del Codice privacy*, in *Dir. Pen. Proc.*, no. 2, 2019;
- MATTIOLI M., *Discolsing Big Data*, in *Minn, L. Rev.*, 2014;
- MANNOZZI G., *Razionalità e “giustizia” nella commisurazione della pena. Il Just Desert Model e la riforma del Sentencing nordamericano*, in *Pubblicazioni della Università di Pavia. Studi nelle scienze giuridiche e sociali*, no. 78/1996;
- MARINI G., *La capacità d’intendere e volere nel sistema penale italiano*, in *Riv. It. Dir. Proc. Pen.*, 1961;
- MARINUCCI G., *L’analogia e la “punibilità svincolata dalla conformità alla fattispecie penale”*, in *Riv. It. Dir. Pen. Proc.*, 2007;
- MARTUCCI P., *Il contributo del criminologo nel processo penale: un problema ancora aperto*, in *Diritto penale e processo*, 2004;
- MAYSON S. G., *Bias In, Bias Out*, in *Yale law journal*, Vol. 128, No. 8, 2019;
- MASSARO A., *Europeizzazione del diritto penale e razionalizzazione del sistema sanzionatorio: il superamento dei “doppi binari” nazionali nel segno sostanzialistico-funzionale della “materia penale*, in *DPC*, 15 July 2015;

ID., *Sorvegliare, curare e non punire: l'eterna dialettica tra "cura" e "custodia" nel passaggio dagli ospedali psichiatrici giudiziari alle residenze per l'esecuzione delle misure di sicurezza*, in *Riv. It. Med. Leg.*, Vol. 4, 2015;

MASUCCI A., *Atto amministrativo informatico (voce)*, in *Enc. Dir.*, Agg.to, vol. I, Milan, 2007;

MATHIESEN T., *Selective incapacitation revisited*, in *Law Human Behaviour*, 22, (4) 1998;

MAUGERI A. M., *L'uso di algoritmi predittivi per accertare la pericolosità sociale: una sfida tra evidence based practices e tutela dei diritti fondamentali*, in *Arch. Pen.*, No. 1/2021;

ID., *I destinatari delle misure di prevenzione tra irrazionali scelte criminogene e il principio di proporzione*, in *Indice Penale*, 2017;

MAUGERI M. – PINTO DE ALBUQUERQUE P., *La confisca di prevenzione nella tutela costituzionale multilivello: tra istanze di tassatività e ragionevolezza, se ne afferma la natura ripristinatoria (Corte Cost. n. 24/2019)*, in *DPC*, 3, 2019;

MAURUSHAT A. – BENNETT L. - MOSES D. – VAILE D., *Using 'Big' Metadata for Criminal Intelligence: Understanding Limitations and Appropriate*, in *ACM Digital Library*, 2015;

MAZZACUVA F., *L'incidenza della definizione "convenzionale" di pena sulle prospettive di riforma del sistema sanzionatorio*, in *Diritto penale contemporaneo*, in *Riv. Trim. – Dir. Pen. Cont.*, No. 3, 2015;

ID., *La materia penale e il "doppio binario" della Corte europea: le garanzie al di là delle apparenze*, in *Riv. It. Dir. Proc. Pen.*, No. 4, 2013;

MCALLISTER A., *Stranger than Science Fiction: The Rise of A.I. Interrogation in the Dawn of Autonomous Robots and the Need for an Additional Protocol to the U.N. Convention Against Torture*, in *Minnesota Law Review*, 2017;

MENECEUR Y.- BARBARO C., *Intelligenza artificiale e memoria della giustizia: il grande malinteso. Interrogativi su una memoria della giustizia catturata nelle correlazioni dell'intelligenza artificiale*, in *Quest. Giust.*, 2020;

MOBILIO G., *L'intelligenza artificiale e i rischi di una "disruption" della regolamentazione giuridica*, in *Rivista di BioDiritto*, No. 2, 2020;

MOAHAN J., *The clinical prediction of violent behaviour*, in *Crime & Delinquency Issues: A Monograph Series, ADM 81-921*, 134, 1981;

- MONACO L. – PALIERO C. E., *Variazioni in tema di crisi della sanzione*: la diaspora del sistema commisurativo, in *Riv. It. Dir. Proc. Pen.*, 1994;
- MONOHAN J. –SKEEM J. L., *Risk Assessment in Criminal Sentencing*, in *Annual Review Clinical Psychology*, 12:489-513, 2016;
- MORO P., *Algoritmi e pensiero giuridico. Antinomie e interazioni*, in *MediaLaws – Rivista di diritto dei media*, No. 3, 2019;
- MORO VISCONTI R., *L'intelligenza artificiale: modelli di business e profili di valutazione*, in *Diritto Industriale*, No. 5, 2018;
- NAGNI E., *Artificial intelligence. L'innovativo rapporto di (in) compatibilità tra machina sapiens e processo penale*, in *Sistema penale*, 2 July 2021;
- NATALE A., *Introduzione. Una giustizia (im)prevedibile*, in *Questione di Giustizia*, no. 4/2018;
- OCCHIUZZI B., *Algoritmi predittivi: alcune premesse metodologiche*, in *Riv. Trim-Dir. Pen. Cont.*, No. 2, 2019;
- OLESON J. C., *Risk in Sentencing: Constitutionally Suspect Variables and Evidence-Based Sentencing*, in *HeinOnline*, 64 S.M.U. L. Rev. 1329, 2011;
- NUVOLONE P., (voce) *Misure di Prevenzione*, EDD, Milan 1976;
- ID., *Il ruolo del giudice nell'applicazione della pena*, in *Trent'anni di diritto e procedura penale*, II, Padua, 1969;
- OSWALD M. - GRACE J. - URWIN S.- BARNES G., *Algorithmic risk assessment policing models: lessons from the Durham HART model and "Experimental" proportionality*, in *Information & Communications Technology Law*, 2018;
- OTRANTO P., *Decisione amministrativa e digitalizzazione della P.A.*, in *Federalismi.it*, no. 2/2018;
- OZER N., *Amazon, Google, Microsofts Are at Odds on the Dangers of Face Recognition. One of Them Is on the Right Path*, in *American Civil Liberties Union*, January, 2019;
- PAGLIARO A., *Commisurazione della pena e prevenzione generale*, in *Riv. it. dir. proc. pen.*, 1981;
- PALMIRANI M., *Big Data e conoscenza*, in *Rivista di filosofia del diritto*, 1, 2020;

- PALOPOLI N., *Il Progetto Ferri fra la Scuola positiva ed il moderno indirizzo criminale*, in *Scuola Positiva*, 1925;
- PAJNO A. (and oth.), *A.I.: profili giuridici. Intelligenza artificiale: criticità emergenti e sfide per il giurista*, in *Biolaw Journal*, no. 3/2019;
- PANNAIN B. – ALBINO M. – PANNAIN M., *La perizia sulla personalità del reo: evoluzione dottrinarica e normativa. Prospettive nel c.p.p. '88*”, in *Riv. It. Med. Leg.*, 1989;
- PAROLI C.-SELLAROLI V., *Sistema penale e intelligenza artificiale: molte speranze e qualche equivoco*, in *Dir. Pen. Cont.-Riv. Trim.*, No. 6, 2019;
- PATRONI GRIFFI F., *La decisione robotica e il giudice amministrativo*, in *Giust. Amm.*, 2018;
- PATTON D. E., *Executive Director del Federal Defenders of New York, Oversight Hearing on "The Federal Bureau of Prisons and Implementation of The First Step Act"*, 2019;
- PELLICCIA R., *Polizia predittiva: il futuro della prevenzione criminale?*, in *Cyberlaw*, 9 May 2019;
- PELUSO C., (voce) *Misure di sicurezza*, in *Dig. Disc. Pen.*, VIII, Turin, 1994;
- PERIN A., *Standardizzazione, automazione e responsabilità medica. Dalle recenti riforme alla definizione di un modello d'imputazione solidaristico e liberale*, in *Rivista di BioDiritto*, 1, 2019;
- PERRY W. L. – MCINNIS B. - PRICE C. C. – SMITH S. C. – HOLLYWOOD J. S., *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*, Rand Corporation, 2013;
- PIERGALLINI C., *Intelligenza artificiale: da mezzo ad "autore" del reato?*, in *Riv. It. Dir. proc. pen.*, 2004;
- POLIDORO D., *Tecnologie informatiche e procedimento penale: la giustizia penale "messa alla prova" dall'intelligenza artificiale*, in *Arch. Pen.*, No. 3, 2020;
- PONTI G., *La abolizione delle presunzioni di pericolosità sociale*, in *Rivista italiana di medicina legale*, IX, 1987;
- PONTI G. - MERZAGORA BETSOS I., *La abolizione delle presunzioni di pericolosità sociale*, in *Riv.it. Med. Leg.*, IX, 1989;

QUATTROCOLO S., *Per un'intelligenza artificiale utile al diritto penale*, in *Biolaw Journal*, No. 2, 2021;

ID., *Sui rapporti tra pena, prevenzione del reato e prova nell'era dei modelli computazionali psico-criminologici*, in *Mimesisjournal*, No. 1, Vol. 22, 2021

ID., *Intelligenza artificiale e giustizia: nella cornice della Carta Etica europea, gli spunti per un'urgente discussione tra scienze penali e informatiche*, in *Legislazione penale*, 18 December 2018;

ID., *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale "predittiva"*, in *Cass. Pen.*, 2019;

S. QUATTROCOLO, *Equità del processo penale e automated evidence alla luce della Convenzione europea dei diritti dell'uomo*, in *Revista Ítalo-Española de Derecho Procesal*, Vol. 1, 2019;

RAM N., *Innovating Criminal Justice*, in *HeinOnline*, 112 Nw. U. L. Rev. 659, 2018;

RECCHIA N., *Artificial intelligence, Big Data and Automated Decision-Making in Criminal Justice*, in *International Review of Penal Law*, 2021;

ID., *Principio di proporzionalità e scelte di criminalizzazione*, in *Dir. Pen. Proc.*, 2020;

REDDING R. E., *Evidence-Based Sentencing: The Science of Sentencing Policy and Practice*, in *Legal Studies Research Paper Series*, Paper No. 09-41, 2, 2016;

RICCI F., *Nuovi rilievi sul problema della "specificità" della prova giuridica*, in *Riv. Trim. dir. Proc. Civ.* 2000;

RICCIO G., *Ragionando su intelligenza artificiale e diritto penale*, in *Riv. Trim. dir. Pen. Cont.* no. 3, 2019;

RIGUZZI F., *Introduzione all'Intelligenza artificiale*, 11 May 2021;

RISSLAND E. L., *Artificial Intelligence and Law: Stepping Stones to a Model of Legal Reasoning*, in *HeinOnline*, 99 Yale L.J. 1957, 1990;

RIVELLO P., *Perito e perizia*, in *Digesto delle discipline penalistiche*, IX, 1995;

RIZER A. – WATNEY C., *Artificial Intelligence Can Make Our Jail System More Efficient, Equitable, and Just*, in *Texas Review of Law & Politics*, Vol. 23, No. 1, 2019;

RIZZI T. – PERA A., *Balancing Tests As A Tool To Regulate Artificial Intelligence In The Field Of Criminal Law*, in *Special Collection and Artificial Intelligence*, 2021;

ROCCA G. – CANDELLI C. – ROSSETTO I. – CARABELLESE F., *La valutazione psichiatrico forense della pericolosità sociale del sofferente psichico autore di reato*, in *Riv. Med. Leg.*, 2012;

ROMANO A., *Rischi e opportunità del ricorso delle amministrazioni alle predizioni dei big data*, in *Dir. Pubbl*, 2019;

RULLI E., *Giustizia predittiva, intelligenza artificiale e modelli probabilistici. Chi ha paura degli algoritmi*, in *Analisi Giuridica dell’Economia*, December 2018;

ROCCA G. – CANDELLI C. – ROSSETTO I. – CARABELLESE F., *La valutazione psichiatrico forense della pericolosità sociale del sofferente psichico autore di reato: nuove prospettive trainadagine clinica e sistemi attuariali*, in *Rivista Italiana di Medicina Legale (e del Diritto in campo sanitario)*, No. 4, 2012;

SAILOR O. C., *At the Nexus of Neoliberalism, Mass Incarceration, and Scientific Racism: the Conflation of Blackness with Risk in the 21st century*, in *Tapestries: Interwoven voices of vocal and global identities*, Vol. 9, Issue 1, 2020;

SALVATI A., *La pericolosità sociale nell’ordinamento giuridico italiano*, in *Amministrazione in cammino*, 11 May, 2011;

SAMUEL A. L. , *Some Studies in Machine Learning Using the Game of Checkers*, in *IBM Journal of Research and Development*, No. 44, Issue:1.2, 1959;

SANTALUCIA B., (voce) *Pena criminale (diritto romano)*, in *Enc. dir.*, XXXII, Varese, 1982;

SANTANGELO A., *Il trattamento illecito di dati all’indomani del Regolamento privacy. Prime ipotesi applicative*, in *Diritto di internet*, No. 2, 2019;

ID., *La rivoluzione dolce del principio rieducativo tra Roma e Strasburgo*, in *Cass. Pen.*, 2019;

SARTOR G. – ROTOLO A., *Agreement Technologies*, in *Law, Governance and Technology Series* 8, 2013;

SCHIAFFO F., *La pericolosità sociale tra “sottigliezze empiriche” e “spessori normativi: la riforma di cui alla legge n. 81/2014*, in *DPC*, 2014;

SEARLE J. R., *Minds, Brains and Programs*, in *The Behavioural and Brain Science*, Vol. 3, Issue 3, Cambridge, 1980;

SELVAGGI, N., *La violenza istituzionale*, in *Violenza, diritto e giustizia*, Dialoghi, 2017;

ID., *La depenalizzazione e le altre politiche deflattive nelle più recenti iniziative di riforma*, in *Archivio penale*, 2014;

SCHERER M. U., *Regulating Artificial Intelligence systems: risks, challenges, competencies, and strategies*, in *Harvard Journal of Law & Technology*, Vol. 29, No. 2, 2016;

SHERMAN L.W., *Ideas in American Policing*, in *Police Foundation*, July 1998;

SEÑOR M., *Gli algoritmi predittivi nell'amministrazione della giustizia*, in *La Voce dell'Agorà*, no. 27, February 2017;

SIGNORATO S., *Il diritto a decisioni penali non basate esclusivamente su trattamenti automatizzati. Un nuovo diritto derivante dal rispetto della dignità umana*, in *Riv. Dir. Proc.*, 2021;

ID., *Giustizia penale e intelligenza artificiale. Considerazioni in tema di algoritmo predittivo*, in *Riv. dir. proc.*, 2020;

SIMONCINI A., *L'algoritmo incostituzionale: intelligenza artificiale e il futuro della libertà*, in *BioLaw, Journal-Rivista di BioDiritto*, No. 1, 2019;

SLOBOGIN C., *Principles of Risk Assessment: Sentencing and Policing*, in *HeinOnline*, 15 Ohio St. J. Crim. L. 583, 2018;

SORBELLO S., *Banche dati, attività informativa e predittività. La garanzia di un diritto penale del fatto*, in *Riv. trim - Dir. pen. Cont.*, no. 2, 2019;

SOMALVICO M., *Intelligenza artificiale*, in *Scienza&Vita*, no. 8, 1987;

SPASARI M., *Appunti sulla discrezionalità del giudice penale*, in *Riv. it. dir. proc. pen.*, 1976;

SPIEGELHALTHER D. J., *The Future lies in Uncertainty*, in *Science*, Vol. 435, 2014;

STARR S. B., *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, in *HeinOnline*, 66 Stan. L. Rev. 803, 2014;

STEVENSON M., *Assessing Risk Assessment in Action*, *Minnesota Law Review*, vol. 103, No. 58, 2018;

STIMSON C.D., *The First Step Act's Risk and Needs Assessment Program: A Work in Progress*, in *Legal memorandum*, The Heritage Foundation, no. 265, 2020;

TENE O. – POLONETSKY J., *Taming the Golem: Challenges of Ethical Algorithmic Decision-Making*, in *HeinOnline*, 19 N.C. J.L. & Tech. 125, 2017;

TONINI P., *Prova scientifica e contraddittorio*, in *Dir. Pen. Proc.*, 2003;

TRAVERSI A., *Intelligenza artificiale applicata alla giustizia: ci sarà un giudice robot*, in *Quest. di Giust.*, 10 April, 2019;

TRAVIS J., *Reflections on the Reentry Movement*, Cuny Academic Works, Vol, 20, No. 2, 2007;

THOMAS E., *Why Oakland Police Turned Down Predictive Policing*, in *Vice.com*, 28 December 2016;

TILLER L., *A Minority Report: The Unregulated Business of Automating the Criminal Justice System* in *The Business, Entrepreneurship & Tax Law Review's B.E.T.R. White Paper*, March 2019;

TRIPODI A. F., *L'idea di pena nei percorsi tracciati dalle Corti europee in materia di ne bis in idem*, in *Il Quaderno di storia penale e della giustizia "Il castigo" Riflessioni interdisciplinari per un dibattito contemporaneo su giustizia, diritto di punire e pena*, Vol. 3, Macerata, 2021;

ID., *Dal diritto penale della paura alla paura del diritto penale. Punti per una riflessione*, in *La paura. Riflessioni interdisciplinari per un dibattito contemporaneo su violenza, ordine, sicurezza e diritto di punire*, in *Quaderno di storia del penale e della giustizia*, Macerata, 2019;

TUZET G., *L' algoritmo come pastore del giudice? Diritto, tecnologie, prova scientifica*, in *Discrimen*, 19 October 2019;

UBERTIS G., *Intelligenza artificiale, giustizia penale, controllo umano significativo*, in *Sistema penale*, No. 4, 2020;

ID., *Il giudice, la scienza e la prova*, in *Cass. Pen.*, 2011;

UNDRILL G., *The risks of risk assessment. Advances in Psychiatric Treatment*, Vol. 13, 291–297, 2007;

VALSECCHI A., *Algoritmo, discrezionalità amministrativa e discrezionalità del giudice*, in *Riv. Dir. Amministrativo*, 14 September 2020;

VIGANÒ F., *Un'importante pronuncia della Consulta sulla proporzionalità della pena*, in *DPC*, 14 November 2016;

VILLASENOR J. - FOGGO V., *Artificial Intelligence, Due Process and Criminal Sentencing*, in *Michigan State Law Review*, Vol. 2020, No. 2, 2020;

VINCENTI E., *Massimazione e conoscenza della giurisprudenza nell'era digitale*, in *Quest. Giust.*, No. 4/2018;

VIOLA L., *L'intelligenza artificiale nel procedimento e nel processo amministrativo: lo stato dell'arte*, in *Foro amm.*, 2018;

WARREN R. K., *Evidence-Based Sentencing: Are We up to the Task*, in *HeinOnline*, 23 Fed. Sent'g Rep. 153, 2010;

WERTH R., *Risk and punishment: The recent history and uncertain future of actuarial, algorithmic, and "evidence-based" penal techniques*, 10th January, 2019;

WEXLER R., *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, in *Stanford Law Review*, Vol. 70, No. 5, 2018;

WIESZ A. N. (et oths), *Assessing the risk of severe domestic violence. The importance of Survivors' predictions*, in *Journal of Interpersonal Violence*, 2000;

WORMITH J. S., *Automated Offender Risk Assessment: The Next Generation or a Black Hole?*, in *American Society of Criminology*, Vol. 16, No.1, 2017;

ZARA G., *Tra il probabile e il certo. La valutazione dei rischi di violenza e di recidiva criminale*, in *Diritto penale contemporaneo*, 20 May 2016;

ZINGALES D., *Risk assessment: una nuova sfida per la giustizia penale? La pericolosità criminale al vaglio algoritmico delle probabilità nell'esperienza statunitense*, in *DPU*, 9 December 2021;

ZIROLDI A., *Intelligenza artificiale e processo penale tra norme, prassi e prospettive*, in *Quest. Giust.*, 18 October 2019;

ZUDDAS P., *Intelligenza artificiale e discriminazioni*, in *Liber amicorum per Pasquale Costanzo*, 16 March 2020.

Online sources

ALEXANDER M., *The Newest Jim Crow*, published in *The New York Times*, 8th November 2018;

ANDERSON C., *The end of theory: the data deluge makes the scientific method obsolete*, in *wired.com*, 23 June, 2008;

ANGWIN J., *Make algorithms accountable*, *New York Times*, 1st August 2016;

ANSPACH R., *How the First Step Act Got People Out of Prison and Back With Their Families*, December, 2019;

BORSARI R., *Intelligenza artificiale e responsabilità penale: prime considerazioni*, in *Medialaw*, 2019;

CALO R., *Artificial Intelligence Policy: a Primer and Roadmap*, in *University of Bologna Law Review*, 2018;

CAMPBELL J. C., *Violence against women: II. Health consequences of intimate partner violence*, in *PubMed*, May 2002;

CASTELLETTI L.- RIVELLINI G. – STRATICÒ E., *Efficacia predittiva degli strumenti di Violence Risk Assessment e possibili ambiti applicativi nella psichiatria forense e generale italiana. Una revisione della letteratura*, in *Journal of Psychopathology*, 2014;

CATANESI R. – CARABELLESE F. – GRATTAGLIANO I., *Cura e controllo. Come cambia la pericolosità sociale psichiatrica*, in *Journal of psychopathology*, No. 1, 2009;

CERVELLI R., *Machine learning: cos'è e come funziona l'apprendimento automatico*, 9 May 2019;

CHANTLER A. – BROADHURST R., *Social Engineering and Crime Prevention in Cyberspace*, in *SSRM Electronic Journal*, 2008;

CYPHERT A.B., *Reprogramming Recidivism: The First Step Act and Algorithmic Prediction of Risk*, in *Seton Hall Law Review*, Vol. 51, 2020;

ID., *Tinker-ing with Machine Learning: The Legality and Consequences of Online Surveillance of Students*, in *Nevada Law Journal*, Vol. 20, No. 2, 2020;

DEMICHELE M. (et oths)., *The Public Safety Assessment: A Re-Vaudation And Assessment Of Predictive Utility And Differential Prediction By Race And Gender In Kentucky*, 2018;

DIAKOPOULOS N., *Algorithmic Accountability Reporting: On the Investigation of Black Boxes in Tow Center for Digital Journalism, Columbia University*, 10 July, 2014;

DUMBILL E., *What is big data*, in *Big Data Now: current perspectives*, O mMedia, O' Reilly Media: California, 2012;

GOTTFEDSON L.S., *Mainstream Science on Intelligence*, in *Wall Street Journal*, New York, 13 December 1994;

LICATA P., *Predictive Analytics*, in *Digital4 Online*, 30 March 2022;

LIPTAK A., *Sent to Prison by a Software Program's Secret Algorithms*, in *The New York Times*, 1st March, 2017;

LIVNI E., *Nei tribunali del New Jersey è un algoritmo a decidere chi esce su cauzione*, in *Internazionale*, March 2017;

MAFFEO V., *Giustizia predittiva e principi costituzionali*, in *www.i-lex.it*, 2019;

MERKL T. A. – ARZY L., *California's Referendum to Eliminate Cash Bail, Explained*, 2nd October 2020;

MORELLI C., *Giustizia predittiva: il progetto (concreto) della Corte d'appello di Brescia*, April 2019;

NILER E., *Can AI Be a Fair Judge in Court? Estonia Thinks So*, in <https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/>, March 2019;

RYBERG J., *Risk-Based and predictive accuracy*, in *Springer link*, 8 February 2020;

SHAPIRO D. L. – NOE A. M., *Risk Assessment. Origins, Evolution, and Implications for Practice*, 2015;

SIGNORELLI A.D., *Il software italiano che ha cambiato il mondo della polizia predittiva*, in *Wired.it*, 18 May 2019;

STRADELLA E., *La regolazione della Robotica e dell'Intelligenza artificiale: il dibattito, le proposte, le prospettive. Alcuni spunti di riflessione*, in *www.medialaws*, 2019;

TREVISI C., *La regolamentazione in materia di Intelligenza artificiale, robot, automazione: a che punto siamo*, in *Medialaws*, 21 May 2018;

TUZET G., *L'algoritmo come pastore del giudice?.Diritto, tecnologie, prova scientifica*, in *Medialaws*, 16th March, 2020;

ZAVRSNIK A., *Criminal justice, artificial intelligence systems, and human rights*, in *Springerlink*, 20 february 2020.

Reports

BRUNDAGE M.– AVIN S.- CLARK J. – TONER H.– ECKERSLEY P.– GARFINKEL B - DAFOE A. – SCHARRE P. – ZEITZOFF T. – FILAR B.– ANDERSON H– ROFF H.– ALLEN G. C.– STEINHARDT J. - FLYNN C. - Ó HÉIGEARTAIGH S. – BEARD S. – BELFIELD H. – FARQUHAR S. – LYLE C. – CROOTOFR. - EVANS O. –PAGE M. –BRYSON J. – YAMPOLSKIY R. – AMODEI D., *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*, February 2018;

CRAGLIA M., *Artificial Intelligence: a European Perspective*. *EU Publication Office*, Luxembourg, 2018;

DESMARAIS S. – SINGH J.P, *Risk Assessment Instruments Validated and Implemented in Correctional Settings in the United States 2*, in *The Council of State. Governments (CSG) Justice Center*, 2013;

FAZEL S, *Prediction of violent reoffending in prisoners and individuals on probation: a Dutch validation study*, 29th January 2019;

HARVARD LAW REVIEW ASSOCIATION, *Selective Incapacitation: Reducing Crime Through Predictions of Recidivism*, in *Harvard Law Review*, 1982;

SALVANESCHI L., *Diritto giurisprudenziale e prevedibilità delle decisioni: ossimoro o binomio*, Report to the 11th Civilian Observer Assembly, 2016;

NIEDERMAN A. S. (et oth), *The Institutional Life of Algorithmic Risk Assessment*, UCLA School of Law, 2019;

RODOTÀ S. – CAPURRO E. R., *European group on ethics in science and new technologies*, Ethical Aspects of ICT Implantés, Human Body, Bruxelles, no. 20, 2020;

SOUTHERLAND V., *With A.I. and Criminal Justice, the Devils in the Data*, in *American Civil Liberties Union*, April, 2018;

TURNER S. (et oths), *Development of the California Static Risk assessment (CSRA): Recidivism Risk Prediction in the California Department of Corrections and Rehabilitation*, Center for evidence-based corrections, *University of California-Irvine*, 2013.

Papers and Conference Presentations

ALLEGREZZA S., *Artificial intelligence and Sentencing in Criminal justice*, Brussels, 27 January 2021;

CAMPBELL J. C., *Risk assessment for intimate partner femicide. What practitioners need to know*, Paper presented at the International Conference on Children Exposed to Domestic Violence, London, 2001;

GUTHRIE C. – RACHLINSKI J. J. – WISTRICH A. J., *Blinking on the Bench: How Judges Decide Cases*, Cornell Law Faculty Publications, Paper 917, 2007;

LAVARINI B., *Neuroscienze e processo penale. Relazione ad un incontro di studio seminario specialistico presso l'Ordine Avvocati di Pinerolo*, October 2012;

MASTROBUONI G., *Crime is Terribly Revealing: Information Technology and Police Productivity*, in *Review of Economy Studies*, 2017;

TERENGI I., *Sistemi decisionali automatizzati e tutela dei diritti: tra carenza di trasparenza ed esigenze di bilanciamento*, Trento, 2021.

Doctoral Theses

KREMER J., *The end of freedom in public places? Privacy problems arising from surveillance of the European public space*, University of Helsinki, 2017;

ROMANO S., *Il ruolo delle prognosi nel sistema sanzionatorio*, University of Milan, 2018.