ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

SCIENZE MEDICHE GENERALI E SCIENZE DEI SERVIZI

Ciclo XXXVI

Settore Concorsuale: 06/M1

Settore Scientifico Disciplinare: MED/42

Advancing Digital Healthcare Research and Machine Learning-Aided Knowledge Extraction amidst the COVID-19 Pandemic

Presentata da: ***Davide Golinelli***

Coordinatore Dottorato

Prof.ssa Susi Pelotti

Supervisore

Prof.ssa Paola Rucci

Esame finale anno 2023

CONTENTS

1. Introduction

This thesis presents the pivotal scientific works carried out during my PhD in the framework of the COKE Project ("COVID-19 Knowledge Extraction framework for next-generation discovery science"), catalysed by the formidable challenge of the COVID-19 pandemic. This collaborative research project involved the Department of Biomedical and Neuromotor Sciences of the University of Bologna, and the STLab, Institute for Cognitive Sciences and Technologies of the Italian National Council of Research, CNR, and it was funded by the Italian National Ministry of University and Education (MUR) in 2022.

The goals of this thesis are threefold:

1. To describe the context of the pandemic and the impetus towards digital healthcare[1,2].

2. To underscore the importance of digitalization in healthcare and research for combating the pandemic[3].

3. To outline the utilization of Machine learning in assisting semi-automatic knowledge extraction and conduction of systematic reviews of scientific literature [4].

This thesis delves into the swift transitional phase from traditional healthcare and research to digital healthcare and research, which refers

to the use of information and communications technologies in medicine to manage illnesses and health risks, and to exploit the most the potential of data in scientific research.

The structure of the thesis is as follows:

Chapter 2 offers an account of how the COVID-19 pandemic has encouraged the adoption of digital technologies in healthcare, presenting the results of a systematic review of early COVID-19 scientific literature.

Chapter 3 presents the rationale behind the COKE Project and illustrates its methods and preliminary results in detail.

Chapter 4 discusses the pros and cons of the digital health transition on scientific research.

2. How the COVID-19 pandemic has favoured the adoption of digital technologies in healthcare research.

The COVID-19 pandemic, much like all global crises in human history, prompted significant health and economic upheavals in numerous nations. Simultaneously, this exceptional circumstance catalysed a shift towards digital alternatives across various industries and society at large. One example of this transformation was evident in education[5]; the entire sector, spanning from elementary schools to universities, formulated new strategies for remote instruction, transitioning from traditional classroom lectures to live conferencing or web-based teaching[6]. Analogously, healthcare providers reacted to the COVID-19 crisis through the swift adoption of digital solutions and advanced technological tools. Amid the pandemic, digital technology can alleviate, or even remedy, many challenges, thereby enhancing healthcare delivery. Digital tools have been employed to tackle acute needs stemming directly or indirectly from the pandemic (e.g., apps for patient tracking, remote triage emergency services). Nevertheless, several solutions developed and implemented during the COVID-19 crisis could be consolidated for

future use, contributing to the delineation and adoption of novel digital care models.

The development of new digital solutions is expanding swiftly[7]. Beyond "video consultations," these alternatives encompass emails, mobile apps, wearable devices, chatbots, AI-powered diagnostic instruments, voice-interface systems, and mobile sensors like smartwatches, oxygen monitors, or thermometers. A novel service category was the supervision of individuals in home quarantine and large-scale population monitoring. Telemedicine and remote consultation have already demonstrated their efficacy during a period when access to health services for patients not afflicted with COVID-19 or those with nonacute COVID-19 conditions is hindered or delayed. Indeed, according to Keesara et al.[8], instead of relying on a historically established model of in-person interactions between patients and clinicians through a face-to-face care model, contemporary health care services and patient assistance can be ensured remotely through digital technologies.

Prior to the COVID-19 outbreak, it was predicted that the digital transformation in health care would be as disruptive as the changes

observed in other industries. However, as discussed by Hermann et al.[9] and confirmed by Perakslis[10], "despite the constant introduction of new technologies, this change had yet to materialize." The propagation of COVID-19 seems to have finally provided an irrefutably compelling reason to wholeheartedly embrace digital transformation. Furthermore, at the time of writing, many nations already faced multiple waves of contagion, and new lockdowns were enforced[11]. As such, it had become imperative to reevaluate the digital technologies that were employed during the emergency phase and consider their continued or cyclical use in the face of potential recurring outbreaks.

According to Hermann et al.[9], digital technologies can be categorized based on the patient needs they address in health care: diagnosis, prevention, treatment, adherence, lifestyle, and patient engagement. Therefore, it was necessary to understand which digital technologies have been adopted to face the COVID-19 crisis and whether and how they can still be useful after the emergency phase. To achieve this, it was crucial to cover as many aspects as possible of digital technology use in health care in response to the COVID-19 pandemic.

In a systematic literature review[1] we followed the Preferred Reporting Items for Systematic Reviews (PRISMA) approach[12], to include

quantitative and qualitative studies using diverse designs to describe which digital solutions have been reported to respond and mitigate the effects of the COVID-19 pandemic. The review focused on health research, which includes biomedical, epidemiological, clinical, public health, and health systems research.

The initial search was implemented on May 11, 2020, and was limited to the timespan from January 1 to April 30, 2020. The search query consisted of terms considered adequate by the authors to review the literature on the use of digital technologies in response to COVID-19. We searched the

MEDLINE and MedRxiv databases. We extracted study characteristics such as the paper title, journal, publication date, type of technology, and patient needs addressed. We categorized the retrieved papers according to patient needs (diagnosis, prevention, treatment, adherence, lifestyle, and patient engagement). For the categorization of patient needs, we adapted the definition by Hermann et al.[9], which reports the concept of "customer needs addressed" by the health care industry, to identify the patient health needs addressed by digital technology during the early phase of the COVID-19 pandemic.

The definition of patient needs is reported in Table 1.

Table 1. Definition of the healthcare needs addressed by digital technologies.

Healthcare needs addressed	Definition
Diagnosis	*"The process of determining which disease or condition explains a person's symptoms and signs."* ["Making a diagnosis", John P. Langlois, Chapter 10 in Fundamentals of clinical practice (2002). Mark B. Mengel, Warren Lee Holleman, Scott A. Fields. 2nd edition. p. 198. ISBN 0-306-46692-9]
Surveillance	*"The continuous, systematic collection, analysis and interpretation of health-related data needed for the planning, implementation, and evaluation of public health practice"* [Public health surveillance, World Health Organization. Available from: https://www.who.int/health-topics/]
Prevention	*"Preventing the occurrence of a disease (e.g. by reducing risk factors) or by halting a disease and averting resulting complications after its onset."* [Can Fam Physician. 1974 Nov;20(11):65-8. What is Preventive Medicine? Clarke EA]
Adherence	*"The degree to which a patient correctly follows medical advice."* [World Health Organization (2003). Adherence to long-term therapies: evidence for action. Geneva: World Health Organisation. ISBN 978-92-4-154599-0.]
Treatment	*"The use of an agent, procedure, or regimen, such as a drug, surgery, or exercise, in an attempt to cure or mitigate a disease."* [Drexler M; Institute of Medicine (US). What You Need to Know About Infectious Disease. Washington (DC): National Academies Press (US); 2010. IV, Prevention and Treatment. Available from:

	https://www.ncbi.nlm.nih.gov/books/NBK209704/]
Lifestyle	"Adoption and sustaining behaviors that can improve health and quality of life." [Lianov L, Johnson M. Physician competencies for prescribing lifestyle medicine. JAMA. 2010;304:202-203.]
Patient engagement	*"To actively involve people in their health and health care."* ["Health Policy Brief: Patient Engagement," *Health Affairs*, February 14, 2013.]

We also built a scoring rubric by cross-classifying the patient needs addressed by the technology (or technologies) reported in each article with the type of technology itself. We relied on the report "Assessing the impact of digital transformation of Health Services" by the Expert Panel on Effective Ways of Investigating in Health (EXPH) of the European Commission[13] to classify the types of digital technologies (i.e., AI, big data, chatbots, electronic health records [EHRs], mobile apps, robotics, sensors, telehealth, and telemedicine), integrating it with terms found within the analyzed articles when necessary (i.e., blockchain, Internet of Things [IoT], internet search engines, social media, and mobile tracing). We also extracted information and classified each technology reported by the selected articles according to health care system targets, grade of innovation, and scalability to other geographical areas. To do this, we

also relied on the classifications and definitions reported by the EXPH (Table 2).

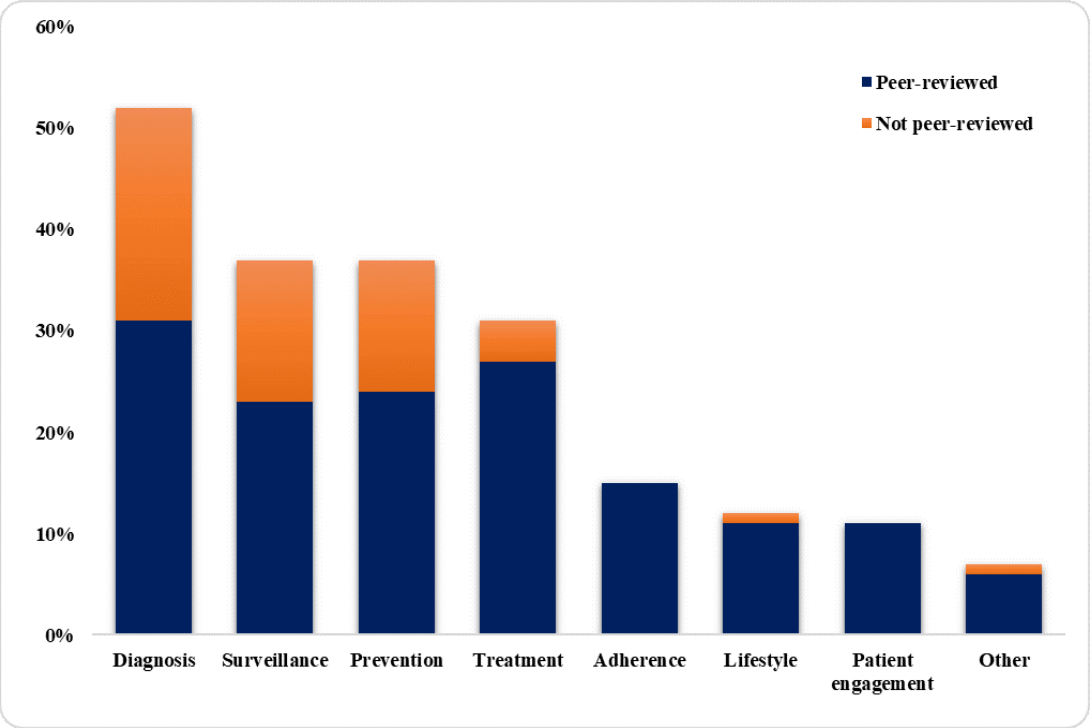
Table 2. Classification of digital technologies and health services.

		Definition
Healthcare system targets	**Clients/patients**	Members of the public who are potential or current users of health services, including caregivers.
	Healthcare providers	Members of the health workforce who deliver health services.
	Health systems/resource managers	Systems/managers involved in the administration and oversight of public health systems. Interventions within this category reflect managerial functions related to supply chain management, health financing, human resource management.
	Data services	Crosscutting functionality to support a wide range of activities related to data collection, management, use and exchange.
Grade of innovation	**Supporting**	Digital services or technologies that can be used to support old or established ones for all or some Healthcare system targets. Such technologies may support or facilitate the performance of existing ones.
	Complementing	Digital services or technologies that can be used in addition to old or established ones for all or some Healthcare system targets. Such technologies may strengthen or enhance the performance of existing ones.
	Substituting	Digital services or technologies that may replace old or established ones for all or some Healthcare system targets.
	Innovating	New digital services or technologies that may offer new possibilities that previously were not available for all or some Healthcare system targets. Such disruptive technologies may represent a new entry into the market.
Scalability to other geographical areas	**Not possible**	Technologies strictly bonded to the context in which they were developed.
	Local	Technologies whose scalability is limited to a local context (i.e. regional or national context), for normative, legislative, ethical or technical reasons.
	Global	Technologies that do not present barriers to scalability such as to prevent their possible global adoption.

The search identified 269 articles (174 from PubMed and 95 from medRxiv), of which 124 full-text articles were assessed and included in the review after screening.

Out of the 124 selected articles, 65 (52.4%) addressed the use of digital technologies for diagnosis (Figure 2), 46 (37.1%) addressed surveillance, 46 (37.1%) addressed prevention, 38 (30.6%) addressed treatment, 15 (12.1%) addressed adherence, 12 (9.7%) addressed lifestyle, 11 (8.9%) addressed patient engagement, and 6 (4.8%) addressed other purposes.

Figure 2. Frequency (%) of appearance of each patient need within the 124 selected articles and share of peer reviewed articles.



Considering the share of peer-reviewed articles, we found that for diagnosis, 39/65 articles (60%) were peer-reviewed; for surveillance, 29/46 (63%); for prevention, 30/46 (65%); for treatment, 33/38 (87%); for adherence, 15/15 (100%); for lifestyle, 11/12 (92%); for patient engagement, 11/11 (100%); and for other, 5 (83.3%).

Even though the SARS-CoV-2 virus triggered a global pandemic, it also promoted the brisk adoption of digital solutions and sophisticated technological tools in healthcare. On one side, medical practitioners and health systems required to monitor large patient populations daily for surveillance objectives[8]. Conversely, they needed swift diagnostic tests for COVID-19 screening, to diminish the workload, and to facilitate patients receiving prompt diagnoses and timely treatments. Such objectives could also be accomplished with the assistance of digital technologies, which were already operative in diverse industries prior to the prevailing crisis. These tools were then hastily deployed in health care as a response to the pandemic[14].

In a systematic review[1] of early scientific literature in reaction to COVID-19, we described numerous digital solutions and technologies addressing a variety of patient and health care requirements. The constantly updated scientific literature serves as a source of significant concepts and recommendations for identifying innovative solutions that ensure patient care during and potentially following the COVID-19 crisis. In the realm of diagnosis, digital solutions that amalgamate with traditional methods of clinical, molecular, or serological diagnosis, like AI-based

diagnostic algorithms based both on imaging and clinical data, appeared promising.

Regarding surveillance, digital applications have already demonstrated their effectiveness[15]; nevertheless, issues related to privacy and usability persisted[16]. To cater to other patient needs, multiple solutions were suggested, including telemedicine or telehealth tools. Although these tools have been available for a considerable period, this historic juncture could potentially facilitate their conclusive large-scale implementation.

The observation that the digital technologies proposed in the analyzed scientific literature primarily address the areas of diagnosis, prevention, and surveillance likely mirrors the emergency phase of the COVID-19 pandemic. As time progressed, well-known digital tools could be proposed for different purposes and patient needs, such as adherence, lifestyle, and patient engagement, which are considered vital determinants of patient health[17] despite the lesser attention afforded to them in early scientific literature.

Apart from the patient needs attended by digital technologies, our review highlighted the most frequently utilized digital technology tools. Given the preliminary phase of the pandemic and its reflection on the articles included in the review, the technologies that have proven to be more readily and swiftly implementable can also be regarded as the most scalable. Indeed, the speed at which these technologies have been deployed underscores their ease of adoption and manageability in a multitude of different contexts, despite their implementation during a pandemic. Numerous solutions have showcased a technical, economic, regulatory, and usability burden that is sufficiently minimal to permit their swift and effective use, at least during the emergency phase. Among these solutions, we document AI tools for diagnosis, big data analytics, and mobile tracing for surveillance and prevention, as well as telemedicine and telehealth, which have proven to be versatile tools for diagnosis, prevention, and treatment.

Many of the digital technologies rapidly implemented during the emergency phase can also be adopted in the subsequent stages of the pandemic[18]. However, this implementation is easier contemplated than executed. According to the article by Keesara et al, "COVID-19 and Health

Care's Digital Revolution," in the context of the digital leap spurred by the COVID-19 pandemic in the United States (and globally), while private corporations and educational institutions have swiftly transitioned to remote work and videoconferencing, the healthcare system is still trailing in adopting digital solutions.

This lag is primarily attributable to the fact that clinical workflows and economic incentives have been devised for a face-to-face care model which, during this pandemic, contributed to the transmission of the virus to uninfected patients seeking medical care. In addition to the historical health care policies, there are limiting factors to the implementation of tools like telemedicine, including an incomplete legal framework designed to regulate the use of innovative IT systems in health care, as well as an inadequate information and communications technology infrastructure and an obsolete reimbursement and payment structure.

Numerous countries are confronting these regulatory issues: the challenges for digital health have surfaced as a global issue in the public health response to COVID-19 and future outbreaks. Digital tools like telemedicine should indeed be incorporated into international and national guidelines for public health preparedness, alongside the

definition of national regulations and funding frameworks in the context of public health emergencies. To transition to new digital care models, enhancing the digital expertise of healthcare professionals and educating the population are essential considerations. Furthermore, by implementing a data-sharing mechanism, digitally collected and stored data will become a valuable instrument for epidemiological surveillance which, as discussed earlier, is critical in controlling the epidemic spread. Lastly, to describe and assess the impact of digital tools during outbreaks, scientific evaluation frameworks should be defined.

The COVID-19 pandemic is facilitating the deployment of digital solutions with an unprecedented pace and impact. It is therefore advisable to keep record of the ideas and solutions being proposed today to implement best practices and care models tomorrow, and to be prepared for future national and international emergencies. It's worth leveraging the momentum provided by the crisis we experienced to implement at least some of the solutions proposed in the scientific literature, especially in national health systems, which in recent years have proven to be particularly resistant to the digital transition.

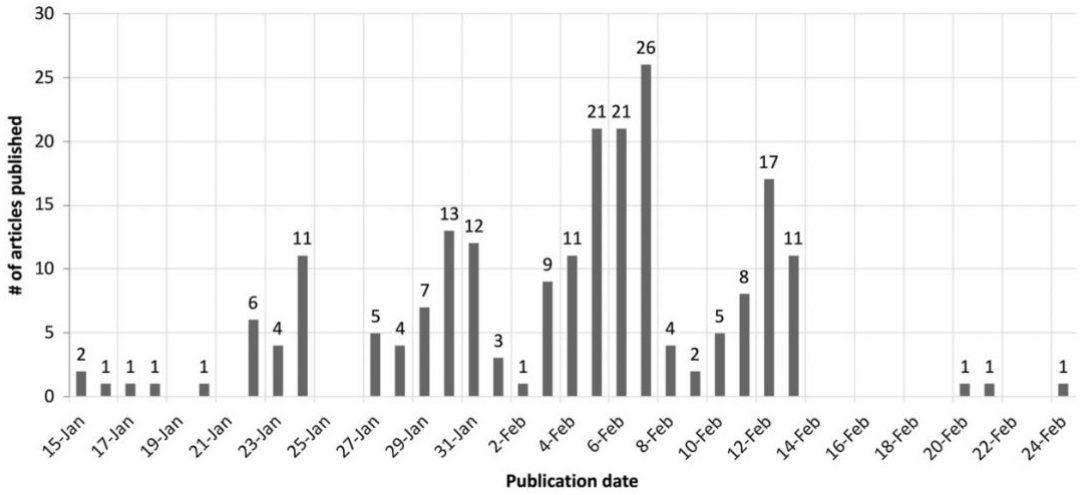
3. The digital health era and its implications for healthcare research: the COKE Project

In the context of rapid digitization of processes and high-speed scientific production highlighted in Chapter 2, the ability of the academic world and healthcare professionals to stay updated and maintain a high level of knowledge is called into question[3]. Moreover, the COVID-19 pandemic itself has further accelerated and amplified this gap between human cognitive abilities to read, extract, and make sense of updated information, and the volume and speed of scientific production.

Figure 3. Representation of the tsunami in scientific literature

Keeping with the paradigmatic example of the pandemic, from January 2020, a continuously growing number of scientific studies related to the novel pathogen emerged in the scientific literature.

Figure 4. Number of papers per publication date in the time-window ranging from January 15th, 2020, to February 24, 2020.



Identifying relevant research outcomes at very early stages is utmost important for guiding the scientific community and governments in more effective research and decisions, respectively. However, traditional methods for measuring the relevance and impact of research outcomes (e.g. citation count, impact factor, etc.) might be ineffective due to the extremely narrow observation window. Notoriously, indicators like citation count or impact factor require broader observation windows (i.e. few years) to be reliable[19]. Altmetrics might be valid tools for measuring the impact in quasi-zero-day time-window. Altmetrics have been

introduced by Priem et al.[20] as the study and use of scholarly impact measures based on activity in online tools and environments.

In a published research[3], we used COVID-19 as a case study for investigating: (1) the tools and frameworks predominantly employed for early scholarly communication; (2) the degree to which Altmetrics could be harnessed to identify potentially impactful research within narrow (i.e. almost instantaneous) time frames. A literature review with stringent eligibility criteria was executed to compile a sample comprised of scientific articles about SARS-CoV-2/COVID-19 appearing in literature within the narrow timeframe spanning from January 15th, 2020 to February 24th, 2020 (figure 4). This sample was utilized to construct a knowledge graph that formally represents the knowledge about papers and indicators.

This knowledge graph fuelled a data analysis process which was employed for experimenting with altmetrics as impact indicators. We detected moderate correlation among traditional citation count, citations on social media, and mentions in news and blogs. Additionally, correlation coefficients were not amplified by indicators associated with zero values, which were common at very early stages following an article's publication. This suggests a shared intended meaning of the

citational acts connected with the aforementioned indicators. Subsequently, we established a method, that is, the Comprehensive Impact Score (CIS), that harmonizes various indicators to provide a multi-dimensional impact indicator. CIS demonstrated promising results as an instrument for selecting relevant papers even within a narrow timeframe. Our findings encourage the advancement of automated frameworks aimed at assisting the scientific community in pinpointing significant work even in instances of limited literature and observation time.

Therefore, our study underscored the disruption caused to scientific production due to the COVID-19 pandemic, and the necessity for innovative digital tools to be utilized to preserve and potentially enhance human ability to navigate through a sea of information that is being produced at a pace often unmanageable by the human mind.

3.1. The COKE Project Rationale

Therefore, the COVID-19 pandemic highlighted the importance of validated and updated scientific information to help policy makers, healthcare professionals, and the public. The speed in disseminating reliable information and the subsequent guidelines and policy implementation are also essential to save as many lives as possible. Trustworthy indications and guidelines should be based on a systematic evidence review of scientific articles which uses reproducible analytical methods to collect secondary data and analyse them. However, the guidelines' drafting process is time consuming and requires a great deal of resources[4].

In general, but even more so in an emergency scenario, it is appropriate to draft and update the guidelines in a relatively short time. This is especially true considering that the scientific community generates an exponentially growing number of scientific papers[21], and that this trend accelerated during the COVID-19 pandemic[22].

Moreover, all of this is taking place in a context in which several new solutions in the management and extraction of information in healthcare are being introduced. For instance, the introduction of information extraction software is an important way of facilitating more

sophisticated healthcare research[23]. Search engines, big data search, and mining tools are continuously being introduced in healthcare processes. The past decade has seen a truly revolutionary paradigm shift to Natural Language Processing (NLP) as a result of which Deep Learning (DL) became the dominating mind-set of researchers and developers in this field[24–26] and became an extremely robust and effective tool for adequately dealing with

the contents of unstructured visual, audio/speech, and textual data.

3.1.1. Standards for Creating Guidelines and Conducting Systematic Literature Reviews

Standard guidelines often address a clinical or policy area, such as COVID-19, and their proper implementation is critical to organizations at the national level, professional associations, healthcare providers, policymakers, patients, and the public[27].

The process of creating, implementing, and evaluating guidelines is extensively depicted in a wide range of literature[28]. Standard guidelines can be quite diverse in terms of scope and focus, ranging from the use of a single medication for a disease or condition, such as the administration of naloxone by laypeople for suspected opioid overdose, to the

comprehensive management of a condition or public health issue, such as the diagnosis, screening, and treatment of type two diabetes mellitus. The formulation of recommendations in a standard guideline can either be entirely new or involve updating existing guidelines. The completion of standard guidelines typically falls within a 9 to 24-month time frame, contingent on their scope. They should be backed by at least one systematic review of the evidence and finalized following one or two meetings of an expert panel. Certain circumstances, such as the COVID-19 pandemic, pose challenges to the creation of standard guidelines. Providing solid evidence in such circumstances is inherently challenging, applicable to both present and potential future pandemics. This is primarily due to the incredibly tight time constraints in emergencies and also due to human cognitive limitations, which play a key role in the process of scientific discovery[29].

Systematic reviews employ reproducible analytical methods to collect and analyze secondary data[30]. They stand apart from traditional literature reviews in that they are more replicable and transparent[31,32]. According to Liberati et al.[32], systematic reviews are crucial tools for summarizing evidence accurately and reliably. They assist clinicians in

staying up-to-date, provide policymakers with evidence to balance the risks, benefits, and detriments of health care behaviors and interventions, compile and summarize relevant research for patients and health care providers, serve as a starting point for clinical practice guideline developers, provide summaries of previous research for funders looking to support new research, and help editors assess the merits of publishing new study reports. This makes them crucial not only for scholars but also for clinicians, researchers, policymakers, journalists, and ultimately, the public.

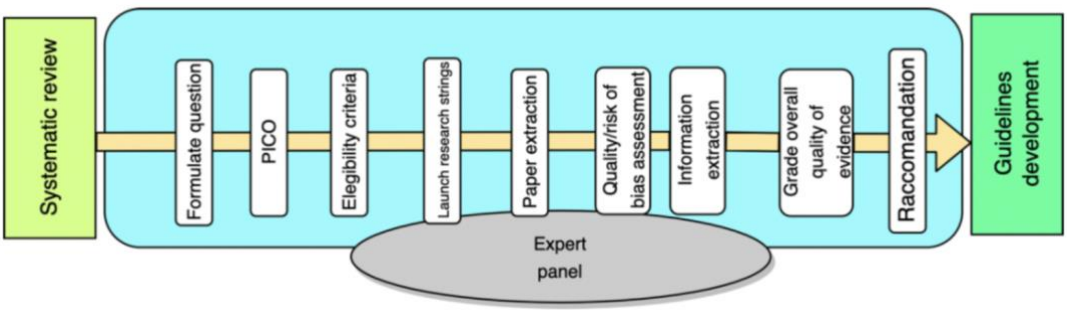
Systematic reviews follow a structured, transparent, and replicable methodology which includes the a priori specification of a research question, clarity on the review's scope and the type of studies eligible for inclusion, comprehensive efforts to locate all relevant research while accounting for possible bias, and analyzing the included studies to draw conclusions based on all the identified research in an impartial and objective manner[33].

Clinical studies and questions always contain four aspects, either explicitly or implicitly: Population/Problem (P), Intervention (I),

Comparison (C), and Outcome (O), known as PICO elements. Utilizing this structure to guide the retrieval of medical evidence within a medical citation database is popular and advantageous. However, accurately and efficiently extracting PICO elements from unstructured information, such as a collection of medical abstracts, can be challenging.

Typically, systematic reviews are carried out by expert panels chosen primarily for their scientific and clinical reputation. Individuals with clear financial conflicts and those whose professional or intellectual bias can diminish the credibility of the review are usually excluded. After defining the research question, the inclusion and exclusion criteria, and the research protocol, the panel experts carry out the operational part of the review process (Figure 5).

Figure 5. Guidelines development steps, including systematic literature review.



The researchers conduct a comprehensive query in multiple databases using search strings and download the records' information (i.e., title, abstract, and other metadata such as the authors' names, journal name, and DOI). The researchers screen titles and abstracts based on the predetermined inclusion and exclusion criteria and the PICO elements. Usually, the majority of the articles are discarded at this stage[30]. Subsequently, the researchers retrieve and read the full text of each individual record included in the new set of articles, and irrelevant articles are discarded. Then, from the final set of articles, the researchers extract the information and evidence relevant to the research question.

Fortunately, the systematic review process involves several explicit and, ideally, reproducible stages, such as identifying all possibly pertinent publications in a uniform way, extracting data from qualified studies, grading, and integrating the findings. These attributes render the process susceptible to automation, which can make it significantly faster, more effective, less prone to errors, and manageable with fewer resources.

A collection of tools have been designed to expedite and semi-automate this process[34,35], including tools to generate terms, visualize and assess search queries, track citation linkages, deduplicate, limit, or translate searches across databases, and prioritize pertinent abstracts for screening. Ongoing studies are leaning towards the creation of tools that can consolidate searching and screening into a single step, and several prototypes of such tools have been successfully developed. A recent scoping review[35] recognized several of these tools (LitSuggest, Rayyan, Abstractr, BIBOT, R software, Robot-Analyst, DistillerSR, ExaCT and NetMetaXL), which hold potential for the automation of systematic reviews. Nevertheless, these tools come with certain limitations, as most algorithms have not yet been transformed into user-friendly tools. While some of these algorithms exhibit high validity and reliability, their application is contingent on user knowledge of computer science and algorithms. In addition, a living systematic literature review[34] examining published methods for data extraction from clinical study reports showed that over 90% of the reviewed publications developed classifiers that predominantly targeted randomized controlled trials (RCTs), while only a handful of tools extracted data from observational studies. These are significant limitations that need to be addressed for the deployment

of automatic or semi-automatic support systems for experts looking to conduct systematic literature reviews on largely unexplored topics, like SARS-CoV-2/COVID-19.

3.1.2. The COKE Project Reasoning and Goal

The necessity for methodological rigor and swift production of sound scientific evidence is of paramount importance in the process of drafting guidelines. This is particularly true during emergencies such as the COVID-19 pandemic, when every decision and its timing can impact the health outcomes of millions of individuals. Therefore, it seems imperative to conceive automated or semi-automated systems to bolster human efforts in the process of screening, extracting, and grading scientific evidence through the deployment of user-friendly tools that can assist the expert panel during the systematic literature review process. Further, during an emergency, it becomes mandatory to include not only the abstract and text from RCTs, but also observational studies, which make up a significant portion of scientific literature.

In light of this, the COKE Project aims to underscore the importance of accelerating and streamlining the extraction and synthesis of scientific evidence in the biomedical domain, specifically within the systematic literature review process. To achieve this, the following part of the thesis is organized into two primary sections. The first (3.2. The COKE Project methods) outlines the framework of the COKE (COVID-19 Knowledge Extraction for next-generation discovery science) Project, which utilizes machine reading and deep learning[36] to semi-automate the systematic literature review workflow in healthcare and, more specifically, outlines our proposed strategy for selecting and navigating relevant literature starting from a set of abstracts obtained by interrogating scientific databases (e.g., EMBASE, Pubmed). The second part (3.3. Preliminary Results: A Case Study on the COVID-19 Literature) presents the project's preliminary results on the automatic classification of sentences into PICO elements in a dataset of abstracts related to COVID-19.

3.2. The COKE Project Methods

The COKE Project, funded by the Italian Ministry for University and Research (MUR) in June 2021 and initiated in November 2021, is a collaborative endeavor between the Italian National Council of Research (CNR) and the University of Bologna. The project's goal is to conceive and implement a semi-automated system that enhances and supports the systematic literature review and guideline drafting processes. Specifically, it aims to expedite the "development" phase (i.e., the systematic review), the "rating/grading" of scientific evidence, and the extraction of pertinent scientific knowledge, with a particular emphasis on the COVID-19 literature.

The COKE Project addresses the following questions: how to automatically link scientific texts or data to cutting-edge knowledge during an emergency or in relation to an emerging disease/pathogen, such as COVID-19? How to adjust and leverage reasoning methods, learning, and reconciling knowledge graphs to recommend serendipitous results to researchers and policy makers?

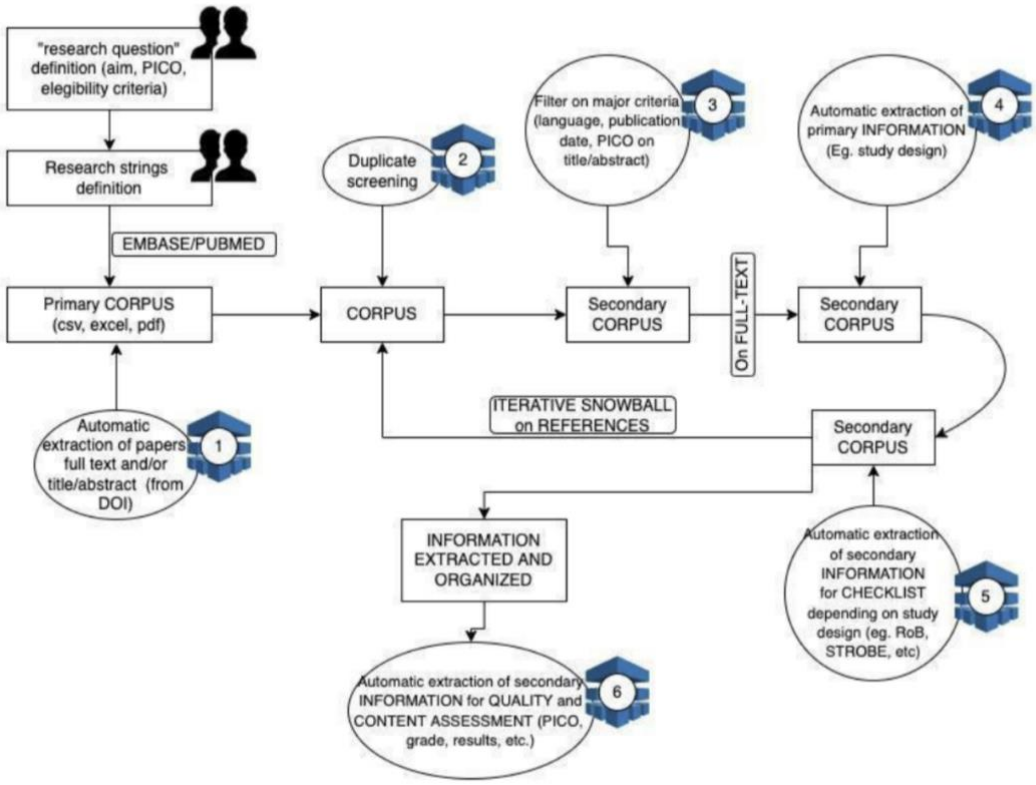
How could a satisfactory level of quality be maintained in knowledge extraction performed by an AI on articles with varying formats and based on diverse study designs when compared to manual research and

content extraction carried out by a domain expert in evidence-based medicine (EBM)?

The COKE Project was structured into four sequential phases: (i) the identification of potentially automatable "nodes" in the workflow of drafting guidelines, specifically in the systematic literature review process; (ii) the creation of a semi-automated system to expedite the process; (iii) semi-automated system trials on research questions connected to the COVID-19 pandemic; and (iv) benchmark tests (human vs machine-aided human). At the moment of drafting, phases (iii) and (iv) of the COKE Project are still in progress.

The guidelines' drafting process is summarized in the previous Figure and Section. The systematic literature review workflow has been refined over the years and involves a series of standardized steps that are already reported in previous section and following Figure 6.

Figure 6. Systematic literature review workflow. The blue icons highlight the potentially automatable nodes of the evidence extraction process.



In the initial stage of a systematic literature review, the investigators establish the research question, the review's objective, and the PICO and eligibility criteria. This preparatory stage is predominantly driven by human involvement. Subsequently, the investigators craft and set the research strings, optimized for each search engine of the chosen platforms/repository/libraries (e.g., MEDLINE, Embase). These queries generate the primary list of records to be screened and potentially analysed. At this juncture, it is possible to automate both the process of

full-text retrieval and the process of eliminating duplicate records (nodes 1 and 2, Figure 6). The full-text retrieval is facilitated by querying the Search API provided by Scopus. The deduplication process is executed by relying on DOIs associated with papers as unique identifiers. In instances where an article does not have a DOI, our disambiguation approach utilises the title and list of authors for such a task.

After the first list of records is obtained, the systematic review workflow employed a filter on several major criteria typically available in the metadata, such as language, publication date, and, most importantly, the PICO criteria. This screening process is normally conducted by the researchers analyzing the title and abstract of each article, and not the full text in its entirety. This decision is taken to ensure high sensitivity and to conserve time without over-utilizing resources. Most records are excluded in the title and abstract phase. Generally, only a small fraction of the records are pertinent, rendering title and abstract screening a significant bottleneck in the systematic reviewing process (node 3, Figure 6)[37]. On one side, the filtering based on metadata is fully automated.

Indeed, COKE performs this task by querying the Scopus API to retrieve all the necessary records that enabled the selection. Among the various metadata, we also gather article-level metrics, such as h-index and altmetrics. On the flip side, the selection based on PICO criteria was semi-automated. This was because an initial step allowed COKE to automatically identify candidate sentences in the title and the abstract of an article that evoked one or more of the PICO criteria, i.e., Participants/Problem (P), Intervention (I), Comparison (C), and Outcome (O). These sentences were then annotated and presented to the expert for a manual assessment. We relied on the automatic annotation of sentences on a deep learning classifier based on three layers, namely BERT, a bi-LSTM layer, and a Conditional Random Field module.

Following the title and abstract screening process, the researchers usually execute the information-extraction process (i.e., study design, study quality, and content analysis). During this phase, the researchers use several design-specific checklists (e.g., RoB 2 tool for randomized clinical trials, or ROBINS-I tool for non-randomized studies) to assess each study's risk of bias. This stage (node 5, Figure 6) could be automated by re-utilizing solutions such as the RobotReviewer system[38,39] that relied

on a Convolutional Neural Network (CNN) architecture trained on corpora derived from the Cochrane Database of Systematic Reviews (CDSR) for rating articles as having a “low”, “high”, or “unclear” risk of bias.

Ultimately, all the necessary information (i.e., PICO-related information, information relevant for the GRADE quality assessment of each study, and other information of interest) were extracted and organized. In this final phase (node 6, Figure 2), we constructed a knowledge graph that organized all the pertinent knowledge extracted with selected articles through previous steps.

The final step, commonly referred to as the "snowball", was executed manually by searching for relevant articles among the references of the studies included in the systematic literature review. This iterative phase was also automatable by using the entries in the reference lists of selected articles as seeds for querying Scopus again and by iteratively restarting our process from the initial step.

Once the aforementioned steps were completed and the necessary information had been obtained, we proceeded with the (manual/human) analysis of the evidence and the drafting of recommendations for the guidelines.

COKE was tested by executing and comparing a series of tasks in a "double-blind" controlled trial between the tool and a team of researchers. In more detail, a COVID-19 research question was formulated. Both the COKE tool and the human researchers reviewed the existing scientific literature and attempted to generate a satisfactory answer.

The outputs were compared in terms of various metrics, such as: the number of papers identified, the quality assigned to each paper analysed, the interpretation of scientific evidence, and results and recommendations.

To guarantee the quality of the COKE procedures, the output was compared to human performance in terms of the classification of study quality and the interpretation of study contents. Indeed, benchmark testing was vital to understand the real-world performance of any

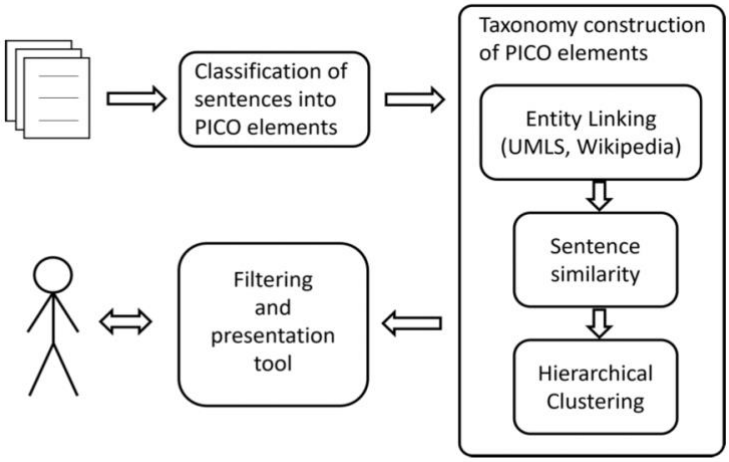
machine-learning-aided system, but such benchmark options were mostly lacking at the time[40].

3.2.1 Organizing and Streamlining Appropriate Literature Based on PICO Elements

A vital facet of our framework dealt with the structuring of the literature to facilitate swift and user-friendly sifting of pertinent articles (node 3, Figure 6).

As previously mentioned, the medical literature has pinpointed four critical aspects in clinical studies, known as PICO elements. Since such information is often absent in a structured format and is described in the abstract text, arranging the obtained literature based on PICO elements demanded a machine comprehension of such text. We then outlined a framework for the structuring and streamlining task based on PICO elements. A broad description is represented in Figure 8.

Figure 8. The proposed framework for organizing and filtering relevant literature based on PICO elements.



Abstracts were initially scrutinized to categorize each sentence into PICO elements (for instance, the sentence “Oral co-amoxiclav (@ mg/kg/day in three doses for @ days) or parenteral ceftriaxone (@ mg/kg/day in a single parenteral dose) for three days, followed by oral co-amoxiclav (@ mg/kg/day in three divided doses for seven days)” for abstract #1319 was identified as I, i.e., Intervention). We then assembled a taxonomy for each PICO element with the objective of structuring the abstracts and enabling filtering. This stage was executed separately for each PICO element after amalgamating related sentences. It comprised three sub-steps. Clinical terms stated in the text were first linked to reference

ontologies and vocabularies of clinical terms (for instance, the unified medical language system (UMLS)[41], SNOMED CT[42], etc.) and other resources (for example, Wikidata[43]), to minimize ambiguity. The similarity of corresponding sentences across abstracts was then computed and hierarchical clustering was executed. Ultimately, the results were transmitted to a user presentation tool for filtering and exploration. We discussed each of the main steps in the remainder of this section.

3.2.2 Sentence Categorization into PICO Elements

Automated categorization of the abstracts' sentences was the initial step to address the aforementioned challenge. To identify PICO elements in text we utilized a linguistic model[44] comprising three principal modules: sentence encoder, sentence contextualization and label sequence optimization layer.

Given a sentence, the first module generated a vector from the sequence of tokens that composed it. To accomplish this task, the authors employed a pre-trained language model named BERT[45]. Pre-trained on English Wikipedia and BooksCorpus, the model was fine-tuned on a large

corpus that merges PubMed abstracts and PubMed Central (PMC) full-text articles in an unsupervised way[46], with the objective to learn the lexicon relevant to the medical domain. The pre-training phase was followed by a supervised learning phase conducted on targeted downstream datasets.

The encoded sentences were subsequently processed by a bi-LSTM layer that contextualized each vector with information taken from adjacent sentences. The output acquired was processed by a feed-forward neural network with only one hidden layer which returned the probability that a sentence belongs to each label.

Lastly, the results acquired from the prediction were processed through a Conditional Random Field (CRF) module that optimized the sequence of labels by modelling the dependencies between them.

3.2.3. Extraction of PICO Elements Taxonomies

Once sentences were categorized, the text related to each PICO element could be extracted and utilized to structure the abstracts accordingly. For instance, a population-based structure would differentiate clinical studies on men from those on women or on children. The user might be

interested in a finer partitioning, e.g patient with a specific disease, or people in an age range. Therefore, we proposed to hierarchically structure the abstracts according to each of the PICO elements extracted from the abstract text. To procure meaningful results, it was essential that the similarity function emphasized the appropriate text features that refer to the population (or the specific PICO element) and disregard unrelated parts.

We proposed to execute entity linking[23,47] to medical and general resources (for example, UMLS, SNOMED CT, MeSH, Wikipedia) to (i) reduce the ambiguity of mentions, (ii) identify the part of the text that are more relevant for each PICO element, (iii) gather hierarchical information from external relevant ontologies and vocabularies. We planned to employ available general, as well as specific entity linking tools, trained on annotated biomedical corpora[48]. We planned to develop a similarity function between entity linked text portions by combining semantic text similarity (for example, Sentence-BERT) and similarity among linked entities. Finally, abstracts were hierarchically organized by clustering[49]. The result of these activities was a knowledge graph (KG) that provided structured knowledge about sentences

collected from abstracts with respect to PICO elements as well as clinical terms.

3.2.4. Filtering Tool

The aim of this component was to reduce the cognitive load associated with the analysis of scientific literature for executing systematic reviews by panels of experts. This aim was achieved in our framework by providing panels with a graphical solution that enabled exploratory capabilities for interacting with the KG. A systematic review process could be classified as an exploratory search task. Exploratory search is extremely time consuming and cognitively complex as it is typically associated with undefined and uncertain goals[50]. For example, there is massive literature associated with the “transmission of COVID-19” and it is inherently hard to identify evidence that can be used for defining guidelines for preventing COVID-19 spread. Hence, in our approach KG exploration supported domain experts in browsing the literature and selecting relevant research works only.

We remind that selected works were then used by experts for finalizing a systematic review process. Filtering was performed by allowing a panel of domain experts to interact with the knowledge graph visually. Accordingly, the filtering tool allowed a panel to explore the literature organized in the graph with respect to PICO elements and clinical terms. This allowed the panel to select one or more concepts in the graph.

In the next section we report the preliminary test's results of the COKE Project related to the COVID-19 abstracts automatic classification.

3.3. Preliminary Results: A Case Study on the COVID-19 Literature

The results of a case study on COVID-19 literature, evaluating the sentence classifier for PICO elements outlined in previous sections, are presented. The aim was to gauge the tool's effectiveness on a distinct topic (i.e. COVID-19), differing from the authors' dataset. We utilized the model honed on the authors' dataset[44,51], a corpus encompassing 24,668 abstracts of randomized controlled trials sourced from Embase. The dataset's annotation was executed automatically via a keyword detection approach. This method accelerates the handling of voluminous data, thereby enabling swifter validation than manual processing. Thus, we evaluated the model's adaptability, trained on such a dataset, to a specific untrained topic.

We examined abstracts related to the pre-exposure prophylaxis of COVID-19, published within a nine-month period on Embase, a widely used, free-access database of medical articles. To facilitate manual validation, 50 abstracts were randomly chosen and segmented into sentences, yielding a total of 752 sentences. These sentences were then annotated by a pair of domain experts. Following[44], seven labels were contemplated: Aim (A), Participants (P), Intervention (I), Outcome (O),

Method (M), Results (R), and Conclusion (C), wherein the Comparison was subsumed into the Intervention category. As the labels Results (R) and Outcome (O) share similar meanings and pose a challenge for experts to differentiate, they were consolidated into a single label Outcome, eventually yielding six labels: A, P, I, O, M, and C. Short sentences that corresponded to subsections (i.e., “Results:”, “Conclusions:”) and ambiguous elements were disregarded by the annotators, leaving 584 annotated sentences (for example, the sentence “PrEP-users identified convenience as a key benefit along with access to PrEP with reduced potential for COVID-19 exposure” of abstract #30 was both manually and automatically annotated as R, Results).

Table 3 outlines the results for each label in terms of precision, recall, and F1. Precision calculates the percentage of predictions with a given label that are accurate, while recall depicts the percentage of true sentences with a given label that are correctly anticipated. F1 represents the harmonic mean between precision and recall and is thus deemed a measure of the classifier's overall performance. Each label's support, i.e., the quantity of sentences annotated to that label, is also noted. The overall performances are reported in terms of accuracy, i.e., the

percentage of accurate predictions, macro average, i.e., the arithmetic average of the performance metrics, and weighted average, i.e., the average of the performance metrics weighted by support.

Table 3. Performance of the PICO classification (Section 3.2) on a dataset of 50 Embase abstracts related to the pre-exposure prophylaxis for COVID-19.

Label	Precision	Recall	F1	Support
A	0.884	0.850	0.867	153
C	0.796	0.854	0.824	96
I	0.312	0.625	0.417	8
M	0.750	0.543	0.630	94
O	0.748	0.821	0.782	184
P	0.479	0.469	0.474	49
Accuracy			0.757	584
Macro avg.	0.662	0.694	0.666	584
Weighted avg.	0.763	0.757	0.756	584

The results revealed a 76% overall accuracy of the classifier. Precision, recall, and F1 exceeded or reached 75% for all labels except I, M, and P, which also demonstrated lower support. As anticipated, the performances were slightly inferior to those on the test set of [44]. This can be attributed to two primary factors. Firstly, the annotation criteria of our dataset differed from the training set's, given that our dataset was manually annotated by experts, while the training dataset was automatically annotated by mapping the abstracts' section headings to labels. Secondly, our dataset exhibited unique features as it pertained to a specific untrained topic (COVID-19) and was not confined to randomized clinical trials.

The results indicate that the applied classifier, trained on automatically annotated data, sustained satisfactory performances in predicting actual (expert validated) labels even on abstracts with diverse topics and characteristics.

4. Conclusions

The relentless increase in the number of papers being published about COVID-19 has now climbed into the tens of thousands and the volume shows no signs of diminishing. This surge of information presents a daunting challenge for scientists, policymakers, and medical professionals alike who are tasked with not only identifying pertinent articles relevant to their work, but also with discerning the validity and quality of the scientific evidence being presented. The burden of distilling this sea of information down to digestible and applicable knowledge has become a crucial issue that requires immediate attention and resolution [21,52].

The COKE Project contributed to discovery science by defining a new knowledge extraction layer in the framework of guidelines development that integrated machine reading and the rigorous protocols of EBM. Through the integration of machine reading with the stringent protocols of Evidence-Based Medicine (EBM), the COKE Project introduces a novel layer of knowledge extraction which can be utilized in the development of medical guidelines. Machine reading has emerged as a key component in this project, facilitating the creation of knowledge graphs directly from

the myriad scientific texts related to COVID-19. Knowledge graphs, which are essentially mathematical graphs that encapsulate factual, conceptual, and procedural knowledge, are formulated as triples (subject, predicate, object) and define binary relationships (via predicates) between entities (i.e., subject and object).

In the dynamic field of healthcare, where knowledge and practices are constantly evolving, the need for swift decision-making that is underpinned by robust scientific evidence is paramount. However, human cognitive abilities have inherent limitations, particularly when faced with the task of extracting and processing vast volumes of information. This study highlights a prominent gap in the scientific domain, particularly within biomedical research, and introduces a project aimed at bridging this gap. More specifically, it aspires to accelerate the development of healthcare guidelines, focusing especially on systematic literature reviews, by semi-automating aspects of the workflow, thereby aiding the human-centric process of extracting and analyzing content.

The COKE Project provides a pioneering framework designed to assist in the systematic literature review process. It does so by methodically examining, efficiently organizing, and effectively filtering medical abstracts. The proposed tool is built upon Natural Language Processing (NLP) techniques which are adept at detecting and categorizing PICO elements (Patient, Intervention, Comparison, Outcome) and medical terminologies, and subsequently arranging abstracts based on these classifications. Early results from the application of the tool on PICO element classification of abstract sentences indicate that a BERT + bi-LSTM language model, trained on an automatically compiled dataset, performed admirably when tested on a real case.

The tool we propose is expected to significantly alleviate the effort required to develop medical guidelines. This reduction in workload, and consequent increase in efficiency, is anticipated to have a substantial positive impact, especially in emergency situations where time is a critical factor. This project thus issues a call to action for similar initiatives that aim to enhance and augment the process of information and knowledge extraction in medicine. Such endeavors are especially crucial

in the ongoing battle against the current COVID-19 pandemic and in the preparedness for potential health crises in the future.

References

1. Golinelli D, Boetto E, Carullo G, Nuzzolese AG, Landini MP, Fantini MP. Adoption of Digital Technologies in Health Care During the COVID-19 Pandemic: Systematic Review of Early Scientific Literature. *J Med Internet Res*. 2020;22(11):e22280. doi:10.2196/22280
2. Sanmarchi F, Fanconi C, Golinelli D, Gori D, Hernandez-Boussard T, Capodici A. Predict, diagnose, and treat chronic kidney disease with machine learning: a systematic literature review. *J Nephrol*. 2023;36(4):1101-1117. doi:10.1007/s40620-023-01573-4
3. Boetto E, Fantini MP, Gangemi A, et al. Using altmetrics for detecting impactful research in quasi-zero-day time-windows: the case of COVID-19. *Scientometrics*. 2021;126(2):1189-1215. doi:10.1007/s11192-020-03809-7
4. Golinelli D, Nuzzolese AG, Sanmarchi F, et al. Semi-Automatic Systematic Literature Reviews and Information Extraction of COVID-19 Scientific Evidence: Description and Preliminary Results of the COKE Project. *Information*. 2022;13(3):117. doi:10.3390/info13030117
5. Robbins T, Hudson S, Ray P, et al. COVID-19: A new digital dawn? *Digit Health*. 2020;6:205520762092008. doi:10.1177/2055207620920083
6. Gewin V. Five tips for moving teaching online as COVID-19 takes hold.

Nature. 2020;580(7802):295-296. doi:10.1038/d41586-020-00896-7

7. Ramsetty A, Adams C. Impact of the digital divide in the age of COVID-19. *J Am Med Inform Assoc*. 2020;27(7):1147-1148. doi:10.1093/jamia/ocaa078
8. Keesara S, Jonas A, Schulman K. Covid-19 and Health Care's Digital Revolution. *N Engl J Med*. 2020;382(23):e82. doi:10.1056/NEJMp2005835
9. Herrmann M, Boehme P, Mondritzki T, Ehlers JP, Kavadias S, Truebel H. Digital Transformation and Disruption of the Health Care Sector: Internet-Based Observational Study. *J Med Internet Res*. 2018;20(3):e104. doi:10.2196/jmir.9498
10. Perakslis ED. Strategies for delivering value from digital technology transformation. *Nat Rev Drug Discov*. 2017;16(2):71-72. doi:10.1038/nrd.2016.265
11. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science*. 2020;368(6493):860-868. doi:10.1126/science.abb5793
12. Moher D. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Ann Intern Med*. 2009;151(4):264. doi:10.7326/0003-4819-151-4-200908180-00135

13. European Commission. European Commission. Assessing the Impact of Digital Transformation of Health Services: Report of the Expert Panel on effective ways of investing in Health (EXPH). Published online November 2019.

14. Madurai Elavarasan R, Pugazhendhi R. Restructured society and environment: A review on potential technological strategies to control the COVID-19 pandemic. *Sci Total Environ*. 2020;725:138858. doi:10.1016/j.scitotenv.2020.138858

15. Sweeney Y. Tracking the debate on COVID-19 surveillance tools. *Nat Mach Intell*. 2020;2(6):301-304. doi:10.1038/s42256-020-0194-1

16. Ienca M, Vayena E. On the responsible use of digital data to tackle the COVID-19 pandemic. *Nat Med*. 2020;26(4):463-464. doi:10.1038/s41591-020-0832-5

17. Corrao G, Rea F, Di Martino M, et al. Effectiveness of adherence to recommended clinical examinations of diabetic patients in preventing diabetes-related hospitalizations. *Int J Qual Health Care*. 2019;31(6):464-472. doi:10.1093/intqhc/mzy186

18. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital Health Strategies to Fight COVID-19 Worldwide: Challenges, Recommendations, and a Call for Papers. *J Med Internet Res*.

2020;22(6):e19284. doi:10.2196/19284

19. Lehmann S, Jackson AD, Lautrup BE. A quantitative analysis of indicators of scientific performance. *Scientometrics*. 2008;76(2):369-390. doi:10.1007/s11192-007-1868-8

20. Priem J, Groth P, Taraborelli D. The Altmetrics Collection. Ouzounis CA, ed. *PLoS ONE*. 2012;7(11):e48753. doi:10.1371/journal.pone.0048753

21. Boetto E, Golinelli D, Carullo G, Fantini MP. Frauds in scientific research and how to possibly overcome them. *J Med Ethics*. Published online October 6, 2020:medethics-2020-106639. doi:10.1136/medethics-2020-106639

22. Aviv-Reuven S, Rosenfeld A. Publication patterns' changes due to the COVID-19 pandemic: a longitudinal and short-term scientometric analysis. *Scientometrics*. 2021;126(8):6761-6784. doi:10.1007/s11192-021-04059-x

23. Malmasi S, Hosomura N, Chang LS, Brown CJ, Skentzos S, Turchin A. Extracting Healthcare Quality Information from Unstructured Data. *AMIA Annu Symp Proc AMIA Symp*. 2017;2017:1243-1252.

24. Hahn U, Oleynik M. Medical Information Extraction in the Age of Deep Learning. *Yearb Med Inform*. 2020;29(1):208-220. doi:10.1055/s-

0040-1702001

25. Wang L, Wang J, Wang M, Li Y, Liang Y, Xu D. Using Internet Search Engines to Obtain Medical Information: A Comparative Study. *J Med Internet Res*. 2012;14(3):e74. doi:10.2196/jmir.1943

26. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. *J Big Data*. 2019;6(1):54. doi:10.1186/s40537-019-0217-0

27. Schünemann HJ, Wiercioch W, Etxeandia I, et al. Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. *CMAJ Can Med Assoc J J Assoc Medicale Can*. 2014;186(3):E123-142. doi:10.1503/cmaj.131237

28. World Health Organization. WHO Handbook for Guideline Development, 2nd ed.; World Health Organization: Geneva, Switzerland, 2014.

29. Kitano H. Artificial Intelligence to Win the Nobel Prize and Beyond: Creating the Engine for Scientific Discovery. *AI Mag*. 2016;37(1):39-49. doi:10.1609/aimag.v37i1.2642

30. Khan KS, Kunz R, Kleijnen J, Antes G. Five steps to conducting a systematic review. *J R Soc Med*. 2003;96(3):118-121. doi:10.1177/014107680309600304

31. Cooper H. *Research Synthesis and Meta-Analysis: A Step-by-Step Approach*. SAGE Publications, Inc; 2017. doi:10.4135/9781071878644

32. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol*. 2009;62(10):e1-e34. doi:10.1016/j.jclinepi.2009.06.006

33. Booth A. Searching for qualitative research for inclusion in systematic reviews: a structured methodological review. *Syst Rev*. 2016;5(1):74. doi:10.1186/s13643-016-0249-x

34. Schmidt L, Olorisade BK, McGuinness LA, Thomas J, Higgins JPT. Data extraction methods for systematic review (semi)automation: A living systematic review. *F1000Research*. 2021;10:401. doi:10.12688/f1000research.51117.1

35. Khalil H, Ameen D, Zarnegar A. Tools to support the automation of systematic reviews: a scoping review. *J Clin Epidemiol*. 2022;144:22-42. doi:10.1016/j.jclinepi.2021.12.005

36. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444. doi:10.1038/nature14539

37. Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to conduct systematic reviews of medical interventions

using data from the PROSPERO registry. *BMJ Open*. 2017;7(2):e012545. doi:10.1136/bmjopen-2016-012545

38. Gangemi, A.; Presutti, V.; Recupero, D.R.; Nuzzolese, A.G.; Draicchio, F.; Mongiovì, M. Semantic web machine reading with FRED. Semant. Web 2017, 8, 873–893.

39. Marshall IJ, Kuiper J, Wallace BC. RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials. *J Am Med Inform Assoc*. 2016;23(1):193-201. doi:10.1093/jamia/ocv044

40. van de Schoot R, de Bruin J, Schram R, et al. An open source machine learning framework for efficient and transparent systematic reviews. *Nat Mach Intell*. 2021;3(2):125-133. doi:10.1038/s42256-020-00287-7

41. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res*. 2004;32(Database issue):D267-270. doi:10.1093/nar/gkh061

42. Donnelly K. SNOMED-CT: The advanced terminology and coding system for eHealth. *Stud Health Technol Inform*. 2006;121:279-290.

43. Vrandečić D, Krötzsch M. Wikidata: a free collaborative knowledgebase. *Commun ACM*. 2014;57(10):78-85. doi:10.1145/2629489

44. Jin D, Szolovits P. Advancing PICO element detection in biomedical

text via deep neural networks. *Bioinforma Oxf Engl*. 2020;36(12):3856-3862. doi:10.1093/bioinformatics/btaa256

45. Devlin J, Chang MW, Lee K, Toutanova K. [No title found]. In: *Proceedings of the 2019 Conference of the North*. Association for Computational Linguistics; 2019:4171-4186. doi:10.18653/v1/N19-1423

46. Lee J, Yoon W, Kim S, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Wren J, ed. *Bioinformatics*. 2020;36(4):1234-1240. doi:10.1093/bioinformatics/btz682

47. Raiman J, Raiman O. DeepType: Multilingual Entity Linking by Neural Type System Evolution. *Proc AAAI Conf Artif Intell*. 2018;32(1). doi:10.1609/aaai.v32i1.12008

48. Mohan S, Li D. MedMentions: A Large Biomedical Corpus Annotated with UMLS Concepts. Published online 2019. doi:10.48550/ARXIV.1902.09476

49. Murtagh F, Contreras P. Algorithms for hierarchical clustering: an overview. *WIREs Data Min Knowl Discov*. 2012;2(1):86-97. doi:10.1002/widm.53

50. White RW, Kules B, Bederson B. Exploratory search interfaces: categorization, clustering and beyond: report on the XSI 2005 workshop

at the Human-Computer Interaction Laboratory, University of Maryland. *ACM SIGIR Forum*. 2005;39(2):52-56. doi:10.1145/1113343.1113356

51. Jin D, Szolovits P. PICO Element Detection in Medical Text via Long Short-Term Memory Neural Networks. In: *Proceedings of the BioNLP 2018 Workshop*. Association for Computational Linguistics; 2018:67-75. doi:10.18653/v1/W18-2308

52. Golinelli D, Toscano F, Bucci A, Carullo G. Transferring Health Big Data within the European Legal Framework: What Role for National Healthcare Services? *J Law Med*. 2018;26(2):488-493.