ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

DATA SCIENCE AND COMPUTATION

Ciclo 37

Settore Concorsuale: 02/D1 - FISICA APPLICATA, DIDATTICA E STORIA DELLA FISICA

Settore Scientifico Disciplinare: FIS/08 - DIDATTICA E STORIA DELLA FISICA

RETHINKING METHODOLOGICAL FOUNDATIONS IN SCIENCE EDUCATION RESEARCH IN THE AGE OF ARTIFICIAL INTELLIGENCE AND BIG DATA

Presentata da: Martina Caramaschi

Coordinatore Dottorato

Daniele Bonacorsi

Supervisore

Olivia Levrini

Co-supervisore

Tor Ole Bigton Odden

Esame finale anno 2026

Abstract

Artificial Intelligence (AI) and data science are no longer peripheral instruments in Science Education Research (SER). They offer new ways to organise, analyse, and produce knowledge. In particular, computational approaches to textual data, such as natural language processing, machine learning, and network analysis, provide models for addressing complex textual corpora, opening new possibilities for integrating the sensitivity of qualitative inquiry with the scope of quantitative analysis.

However, the adoption of AI-based methods is not merely a technical innovation but entails a profound shift in methodological assumptions. Two main gaps emerge: these assumptions remain largely underexplored, and we still lack a clear understanding of how AI-based methods relate to existing paradigms in SER. Moreover, the field lacks systematic instruments for analysing such methodological shifts and positioning them within the broader paradigmatic landscape.

To address this gap, this thesis develops a methodological and philosophical reflection on how computational and AI-based approaches interact with qualitative inquiry in SER. Through two case studies on text analysis, it examines how these methods operate in practice and what kinds of knowledge they produce. From these analyses, a heuristic tool was designed to make visible the assumptions that govern research methods, structured along ontological, epistemological, and axiological dimensions, and subsequently applied to AI-driven methods in contemporary scientific research and to emerging hybrid approaches in science education.

The results show that AI does not simply accelerate existing ways of producing scientific knowledge but expands them, acting as an epistemic agent that places data curation and manipulation at the centre of research. Furthermore, the combination of computational and qualitative approaches in SER proves most fruitful when their complementary features are valued within an iterative dialogue. By making these transformations explicit, the thesis contributes to rethinking mixed methods not simply as a procedural combination of qualitative and quantitative approaches, but as a paradigmatic condition in which AI-based representations intrinsically intertwine statistical and interpretive dimensions of data, and to a renewed vision of methodological pluralism in the age of Artificial Intelligence.

This is not just a mapping exercise.
It is the recognition that something fundamental
in the methodological landscape is shifting.
This thesis positions itself inside that shift.

General introduction of the thesis

This thesis is a methodological reflection on how science education research approaches the analysis of textual data, with a specific focus on the changes introduced by computational artificial intelligence-based (AI) methods. The core concept addressed in this thesis is the evolving concept of mixed methods, seen as the possibility of integrating qualitative and quantitative analyses in the same research. The thesis also discusses broadly what changes are introduced by AI-based methods in scientific research, reflecting on how this is changing the nature of methods that students in science classrooms should be aware of.

The initial motivation for this research emerged from a specific case: the need to analyze a medium-sized corpus of about two hundred essays, in which students described their ideal future lives, as part of a project on future-oriented science education. This case, like others in science education, where substantial textual datasets are available to inform research but are too large to be addressed solely through traditional qualitative analyses, raised the question of whether computational methods could be effectively employed for the analysis of such data in educational contexts. At the same time, the idea of abandoning qualitative methods did not seem appropriate, as it allowed researchers to immerse themselves deeply in the data and to listen attentively to the individual voices of students. It was at this point that the need for an approach capable of integrating both qualitative sensitivity and quantitative breadth became evident: the need for mixed methods.

As AI, data science, and computational tools increasingly enter the field of science education, they open up new possibilities for analyzing textual data and for combining qualitative and quantitative approaches. However, these possibilities also raise important theoretical and practical questions, often challenging traditional methodological boundaries. This thesis critically explores these developments, not only in terms of their practical applications, but also by examining the epistemological, ontological, and axiological assumptions that underpin them.

The thesis is organised into five chapters, reflecting the evolution of this methodological research. First, it was essential to engage theoretically with the concept of mixed methods, to understand how qualitative and quantitative approaches have been traditionally combined in educational research and what philosophical assumptions underlie their integration. **Chapter 1** presents a critical review of mixed methods research, both in general education and specifically within science education. A review of key educational research handbooks and major contributions reveals that mixed methods are often framed within a pragmatic paradigm, focusing primarily on procedural aspects. However, such procedural accounts risk overlooking the deeper epistemological, ontological, and axiological assumptions that govern the choice and combination of methods. This kind of oversight becomes even more significant when computational AI-based techniques are involved, whose philosophical foundations are still emerging.

Within this review, particular attention is devoted to three central contributions: Creswell and Guetterman's comprehensive work on educational research methods; the pragmatic vision of mixed methods proposed by Johnson and Onwuegbuzie; and the science-education-centered reflections of Treagust and Won. The latter's work constitutes a turning point in the literature review, offering both conceptual language and critical clarity to articulate the tensions between technical practice and foundational reflection in mixed methods research. Aligning with Treagust and Won's perspective,

this thesis argues for the need to move beyond procedural pragmatism and towards a deeper critical engagement with the philosophical foundations of methodological choices, particularly when AI-based quantitative approaches are considered.

Building upon this theoretical groundwork, **Chapter 2** shifts to direct engagement with computational methods. In order to fully appreciate the affordances and limitations of AI-based techniques in the analysis of textual data, it was necessary to apply these methods firsthand. This phase of the research involved two case studies.

The first case study focuses on the analysis of a dataset of more than 12,000 articles related to physics education, published across sixty years in two major journals. This large-scale thematic analysis allowed me to familiarize myself with natural language processing (NLP), topic modeling, and other computational techniques. It also provided an opportunity to explore their technical possibilities, as well as the creative adaptations and methodological uncertainties that arise when applying them in educational research.

The second case study concerns the dataset that originally motivated the research: a medium-sized corpus of about two hundred student essays on their ideal future lives, within the context of future-oriented science education. In this case study, I have reconstructed and described what happened when both a qualitative analysis using thematic analysis and a computational method based on topic modeling were applied to the same dataset. This dual approach allowed me to compare the strengths and limitations of each method and set the stage for the discussions in **Chapter 3**, where a comparison of the two methods for analyzing educational textual data is provided.

The application of computational tools soon revealed that technical proficiency alone was insufficient. A deeper reflection was needed, not only on how methods are used, but on more subtle, often tacit elements: the nature of the context in which data are collected, the forms of knowledge they employ and generate, and the values they embed. This recognition led to the development of a heuristic tool, operationalised in the form of a grid, designed to systematically examine and reflect on the features and assumptions of different methods. Chapter 3 introduces this heuristic tool, explaining its construction and theoretical basis. The heuristic tool is structured around three main methodological dimensions —ontology, epistemology, and axiology— and includes specific methodological lenses, such as the role of the researcher, the nature of the data, and the types of results produced. Its purpose is to move beyond surface-level descriptions of methods and foster a more critical, explicit awareness of the underlying assumptions driving educational research methods.

Chapters 4 and 5 apply this heuristic tool in two different directions. **Chapter 4** uses the operationalised grid to examine how emerging AI-based methods are being discussed and developed in the broader field of scientific research, particularly considering the methodological and philosophical shifts they entail. This exploration helps position science education within a broader landscape of methodological innovation and debate, while also contributing to a redefinition of the nature of methods adopted in scientific research, which is crucial for reshaping how scientific methods are taught in the classroom.

Chapter 5 then applies the operationalised grid to a selection of recent studies in science education that incorporate computational techniques such as machine learning, network analysis, and NLP. To enrich this analysis, a series of expert interviews were conducted with researchers actively working at the intersection of AI and science education. These interviews, structured using the heuristic tool, enabled a deeper, dialogical exploration of the methodological assumptions underpinning current

practices. The goal is to assess to what extent these studies can be meaningfully understood as mixed methods research and, more ambitiously, to reflect on how the very concept of mixed methods might need to be reformulated in light of the growing presence of AI-based approaches in textual data analysis within science education research.

In essence, this thesis is both a methodological inquiry and a personal intellectual journey. This work mapped paradigms in science education research, where boundaries between different approaches often blur under the advent of digital and emerging technologies. I critically combined qualitative and computational methods for text analysis to gain an insider's view of how knowledge analysis unfolds. In addition, I developed and applied a heuristic tool to make explicit the foundational assumptions that underlie research practices, offering a way to re-examine paradigms in this new technological age. Together, these perspectives, tools, and reflections aimed at fostering a more critical, nuanced, and philosophically grounded understanding of how we conduct research, especially as emerging technologies challenge the traditional foundations of our methods. This reflection is particularly relevant in the field of science education, where methods are not only instruments for research, but also content to be taught, part of the "tools of the trade" that students must learn to fully grasp how science operates. This thesis seeks to convey the complexity and delicacy of research methodologies, an aspect too often taken for granted or overlooked.

To support the reader in navigating the structure of the thesis and appreciating both the specific contributions of each chapter and their broader conceptual interconnections, the following visual map provides an overview of the organisation of the work.

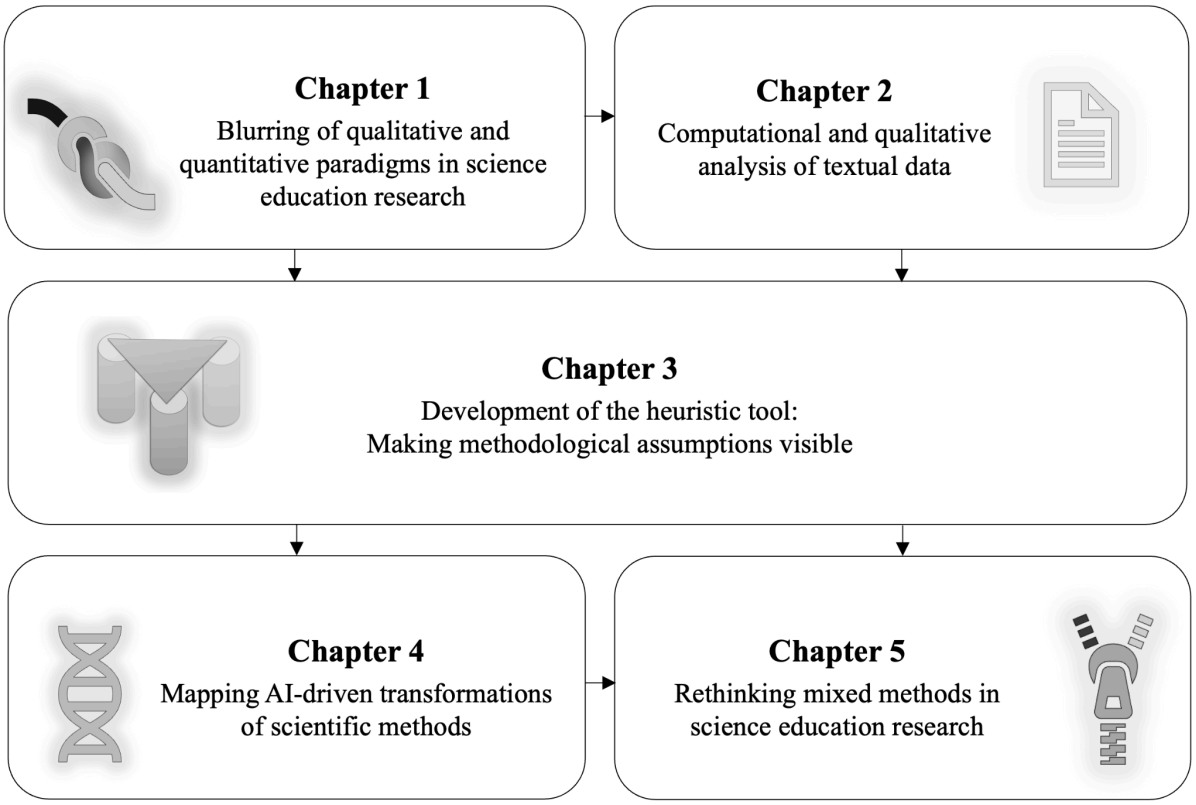

Figure I.1 - Conceptual map of the thesis, illustrating the progression from initial methodological exploration, through the development of the heuristic tool, to its subsequent application.

Along the course of this doctoral research, parts of the studies presented in this dissertation have been published or are currently under review. These publications correspond to specific analyses and methodological developments discussed throughout the chapters of this thesis. For ease of reference, they are listed below:

1. Qualitative analysis of students' narratives in science education:

   Barelli, E., Tasquier, G., Caramaschi, M., Satanassi, S., Fantini, P., Branchetti, L., & Levrini, O. (2022). *Making sense of youth futures narratives: Recognition of emerging tensions in students' imagination of the future.* Frontiers in Education, 7, 1–17. https://doi.org/10.3389/feduc.2022.911052

2. Comparison between qualitative and computational methods applied to the same textual dataset:

   Caramaschi, M., Zanellati, A., & Levrini, O. (2025). *Comparing Textual Data Analysis Methods in Science Education Research.* In *Connecting Science Education with Cultural Heritage.* Springer. https://doi.org/10.1007/978-3-031-90944-3

3. Preliminary large-scale literature review using NLP methods:

   Caramaschi, M., Odden, T. O. B., & Levrini, O. (Accepted, in press). *Reviewing 55 years of Literature on Physics Education using Natural Language Processing.* In *4th World Conference on Physics Education Proceedings.* Springer.

4. Large-scale NLP analysis of physics education corpora related to teaching and learning:

   Caramaschi, M., & Odden, T. O. B. (2025). *Analyzing the history of physics education in the USA and Europe through natural language processing.* Physical Review Physics Education Research, 21, 020153. https://doi.org/10.1103/wfvw-hkyy

5. Analysis of the methodological novelties introduced by AI in scientific research and their implications for science education:

   Caramaschi, M., Levrini, O., & Erduran, S. (Submitted). *Learning from AlphaFold: AI-Driven Methodological Shift in Science for Science Education.* Science & Education.

Index

Chapter 1 – Crossing blurred boundaries

Paradigms and new approaches in Science Education Research

1.1 Introduction

This first chapter positions the thesis within the broader context of science education research and introduces the discussion on mixed methods in this field. It begins with an overview of history and methodological traditions in science education research.

Traditionally, two major paradigms, which are post-positivist and interpretivist, have guided the field (Section 1.2). While these paradigms have provided powerful lenses to study teaching and learning, they reveal their limitations in the face of modern hybrid research contexts. New forms of data, settings and computational tools blur the boundaries between what has long been considered "quantitative" or "qualitative", creating research situations that resist being fully captured by either paradigm. This signals the possibility that we are not merely extending established traditions but are entering a new methodological landscape.

Against this backdrop, the chapter turns to the so-called "third" paradigm of mixed methods research, which may provide conceptual and methodological resources to address these evolving challenges. Thus, we provide a reconstruction of the state of the art of mixed methods research, where relevant literature is critically discussed (Section 1.3). Particular attention is devoted to three major contributions.

First, we show the concept of mixed methods in education as discussed in Creswell and Gutterman's widely cited handbook on research methods (2019). Second, we enter the foundations of mixed methods, examining the vision proposed by Johnson and Onwuegbuzie (2004), who frame mixed methods within a pragmatic maxim. Then, we move to the more science-education-specific contribution recently provided by Treagust and Won (2023). At this stage reasons for critically considering the pragmatic approach are presented, calling for deeper epistemological reflection.

To better understand how these conceptual frameworks have been interpreted and applied in science education research over the past two decades, this section presents a review of the existing literature on mixed methods approaches in the field. Rather than attempting to define a unified perspective, the aim is to provide a broad overview of the diverse voices, case studies, and methodological reflections that have emerged, highlighting both the contributions and the persistent challenges documented so far.

Building on this review, we point out specific challenges and changes (Sections 1.4 and 1.5). One concerns the transformative nature of textual data from text to vectors of numbers, characteristic of the era of big data and data science, which is questioning the division of data in qualitative or quantitative information. Another is the novelty introduced by the adoption of Artificial Intelligence (AI) and computational methods as a quantitative counterpart for qualitative methods. Moreover, attention is devoted to the specificity of educational contexts such as science education.

The chapter concludes by unveiling the underlying attitude of this work (Sections 1.6 and 1.7). It aligns with those contributions that, while engaging with mixed methods, emphasise the importance of reflecting on what is being integrated, how the integration is carried out, and which assumptions

and paradigms inform methodological choices. This reflective stance becomes especially relevant considering the new challenges and possibilities introduced by AI-based methods. These developments invite a renewed discussion on how computational tools can be meaningfully situated within mixed methods research, particularly when working with textual data in science education.

1.2 Research traditions in Science Education Research

1.2.1 Qualitative and quantitative approaches in Science Education Research

Science Education Research (SER) is an academic field devoted to the systematic study of how science is taught, learned, and understood in diverse contexts. One of its primary aims is to study, understand, and generate knowledge to improve the broad educational practice. For instance, investigating students' learning processes, teachers' professional knowledge, and the impact of curricula and pedagogical innovations is an important and frequent endeavor. As a research field, SER is inherently methodological, drawing on a wide range of approaches including theoretical analyses, historical and philosophical perspectives, as well as empirical investigations of teaching and learning.

Within this field, different research orientations are adopted. The two most influential are the so-called qualitative and quantitative. To give an initial idea of their characteristics, the quantitative tradition emphasizes measurement, standardization, and large-scale assessment, aiming to identify generalizable patterns across contexts. From this perspective, educational phenomena are conceptualized as sets of variables whose interactions can be modeled and statistically tested, much as natural phenomena are studied within theoretical frameworks. By contrast, the qualitative tradition seeks to illuminate the complexity of meanings, practices, and settings. It focuses on the context of a phenomenon, assuming that human behavior is inseparable from the circumstances in which it occurs. Rather than reducing social interactions to variables, qualitative studies privilege the situated, interpretive, and often unique nature of events, drawing on methods such as classroom observation, interviews, discourse analysis, and ethnography (Libarkin & Kurdziel, 2002).

Essentially, these traditions reflect two different needs of research: on one side, the need to capture generalizable patterns and stable, reproducible results; on the other, the intention to investigate particular situations dense with situated meanings and to navigate their complexity.

It should be acknowledged, however, that the labels *qualitative* and *quantitative* are themselves controversial, as they may refer variously to data types, methodological strategies, or forms of knowledge produced (Taber, 2014). In the following sections, these two traditions will be clarified by situating them within research paradigms that define elements such as the worldview behind these approaches, that sets the conditions in which methods can operate, depicting the methodologies in a more complex way (Szyjka, 2012; Treagust and Won, 2023).

Having outlined their broad characteristics, it is important to examine in more detail how the qualitative and quantitative orientations conceptualize knowledge and approach research. The quantitative orientation emphasises measurement, impartiality attitude of the researcher, and the search for generalisable findings. It is rooted in the assumption that social facts possess an objective reality which can be accessed independently of the researcher's perspective. Guided by a predominantly deductive logic, it typically begins with hypotheses or theories that are tested through structured designs such as experiments, surveys, or correlational studies (Isaac & Michael, 1995; Glesne, 2006). Variables are narrowly defined and carefully controlled so that the relationships

between them can be revealed through statistical analyses. This approach makes it possible to identify patterns across large populations, to test cause-and-effect relationships under specific conditions, and produce findings that are presented as transferable to wider contexts. Its influence is particularly visible in policy and administrative domains, where numerical indicators and statistical evidence are often treated as the most trustworthy form of knowledge (Feuer, Towne, & Shavelson, 2002). However, its rigidity and abstraction can constrain opportunities to generate theory from lived experience, and findings may reflect trends that risk being disconnected from the immediate realities of classrooms and participants (Johnson & Onwuegbuzie, 2004).

By contrast, the qualitative orientation rests on the assumptions that reality is socially constructed, and variables within any given situation are highly complex, so irreducible to discrete variables (Glesne, 2006). Its purpose is to contextualise and interpret phenomena in-depth, often beginning with inductive inquiry that leads to participant-generated hypotheses or theories. The researcher is considered the main instrument, working in settings that are as naturalistic as possible, and directly engaging with participants in a highly personal way (Rossman & Rallis, 1998). Methodologies such as ethnographies, grounded theory, case studies, phenomenologies, and narratives (Bogdan & Biklen, 2003; Creswell, 2003) allow the investigator to choose the strategy best suited to the research question, often relying on interviews, observations, fieldwork, content analysis, and open-ended surveys. A defining characteristic of qualitative inquiry is its emphasis on "thick description" (Denzin & Lincoln, 2000), which moves beyond surface-level accounts to provide a detailed and nuanced understanding of social phenomena. This richness comes at the cost of time: qualitative research requires sustained engagement, iterative analysis, and careful attention to detail. It is also inherently interpretive, with findings reflecting not only the perspectives of participants but also the values, experiences, and positionality of the researcher. Issues of trustworthiness are therefore paramount, and are evaluated through criteria such as credibility, applicability, consistency, and neutrality (Guba, 1981, cited in Krefting, 1991). Techniques like triangulation, which align multiple data sources or perspectives, help to enhance validity and provide a more holistic portrayal of the phenomenon under study (Anfara et al., 2002). Unlike quantitative research, qualitative inquiry does not claim generalisability in the statistical sense, but instead offers depth, flexibility, and responsiveness to context.

Taken together, these two traditions reflect complementary but distinct visions of knowledge production in science education research. Quantitative methods privilege breadth, predictive power, and generalisability, while qualitative approaches prioritise depth, contextualisation, and meaning.

1.2.2 A shift from quantitative to qualitative approaches

The influence of these orientations has shifted over time. In particular, the focus on qualitative approaches became increasingly significant during the 1960s and 1970s, when the emphasis in SER moved from theory verification and generalised outcomes to understanding phenomena in their situated and specific contexts (Szyjka, 2012). As interests have moved toward the study of classroom experiences and related programs, qualitative text-based data have become a fundamental resource for educational inquiry, providing critical access to understanding learning processes.

A notable example of this evolution can be found in Physics Education Research (PER) (Ding, 2019). In its early developments, PER strongly relied on quantitative techniques to produce and interpret empirical data, reflecting the disciplinary training of many of its pioneers. Researchers with backgrounds in physics naturally gravitated toward methodological frameworks that privilege

objectivity, standardisation, and numerical rigour. Their commitment to scholarly objectivity, combined with technical skill in quantitative analysis, fuelled the popularity of this paradigm in the formative years of PER. Over time, however, the field broadened to incorporate qualitative perspectives, which were necessary to capture the lived realities of classrooms, the nuanced reasoning of students, and the interpretive practices of teachers.

1.2.3 The historical debate on method in social sciences

The division between quantitative and qualitative research has deep philosophical roots. A historical perspective helps to better understand this dualism and also to justify the shift previously described. It originates in the late nineteenth-century debates on the nature of knowledge in the natural versus the human sciences. As Fantini (2014) notes, figures such as Wilhelm Dilthey and Wilhelm Windelband offered contrasting visions: Dilthey distinguished between studying natural phenomena externally and understanding historical-social reality from within, while Windelband proposed the distinction between *nomothetic* sciences, oriented toward universal laws, and *idiographic* sciences, oriented toward the particular. These debates established enduring tensions between positivist and interpretivist orientations, later elaborated by Weber, Simmel, the Frankfurt School, symbolic interactionism, and ethnomethodology.

For education, the significance of this debate lies in the nature of its objects of study: conceptions, learning processes, practices, and identities. These are not "natural facts", but socially embedded constructions that require interpretive frameworks to be meaningfully analysed. Within this intellectual backdrop, two orientations came to dominate SER. The quantitative orientation, strongly shaped by the culture of physics and consolidated through PER, emphasised measurement, standardisation, and large-scale assessment. Tools such as concept inventories, surveys, and statistical analyses promised objectivity, comparability, and cumulative insight across studies (Ding, 2019). By contrast, the qualitative orientation, influenced by the social sciences, foregrounded meaning, interpretation, and context. Its methodologies, such as interviews, classroom discourse analysis, ethnography, grounded theory, enabled the exploration of perspectives, practices, and identities that elude standardisation. As Fantini (2014) emphasises, even within qualitative inquiry, methodological traditions vary: ethnography aims at narrative reconstruction, while grounded theory aspires to build theoretical models from data, often oscillating between positivist aspirations and constructivist sensibilities.

In the following Figure 1.1, the main features of these two methodological traditions are compared, as pointed out by Fantini's third Chapter (2014).

These approaches to research shall be understood as characteristic epistemological and ontological views towards knowledge, with consequently methodological implications (Ding, 2019). The differences in their paradigmatic bases determine the very conditions under which each method can be meaningfully employed. Thus, quantitative and qualitative methods are mobilised to inquire into different aspects of a phenomenon and to address different kinds of research questions (Libarkin & Kurdziel, 2002; Szyjka, 2012).

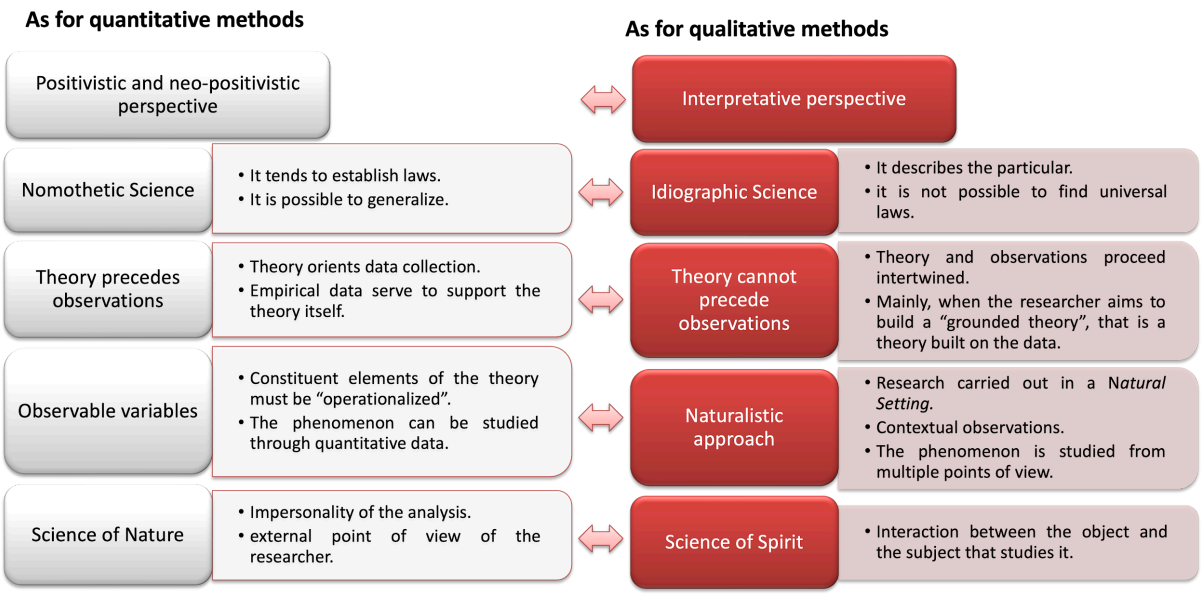

Figure 1.1 - Summary of the foundational features of the two different traditions on social inquiry, neo-positivistic and interpretivist, presented by Fantini (2004), Third chapter.

1.2.4 Convergence and interconnections between the two approaches

The previous sections have outlined two markedly different approaches, representing two methodological poles in social research. However, the scholarly debate has also sought to identify points of dialogue and elements of convergence between them. This section presents some of these shared aspects, with the ultimate aim of clarifying how hybrid approaches have emerged to mediate between the two main traditions.

A crucial point of convergence between quantitative and qualitative methods in SER lies in the nature of what is being investigated (Ding, 2019). In both cases, the reality under scrutiny consists of *second-order constructions*, that are students' learning processes, conceptions, emotions, or instructional practices (Schutz, 1967). These phenomena are not reducible to directly measurable, physical quantities (*first-order constructions*) but are always mediated by interpretation, representation, and context. According to Schutz, in the natural sciences it is possible to develop instruments that directly measure the features under study, for example, determining the duration of a process or combining time and distance to calculate velocity. In contrast, in the social sciences there is no direct access to constructs such as "learning". They can only be approached indirectly, through secondary indicators that reflect how individuals describe, argue, perform, or interact. Consequently, both qualitative and quantitative data in SER are necessarily interpretive, since they rely on these indirect traces to infer the underlying processes. For this reason, quantitative practices in SER should not be crudely regarded as less constructivist than their qualitative counterparts. They also rely on theoretical assumptions, social conventions, and design choices that shape what is measured and how it is interpreted.

Recognising this shared constructivist foundation helps explain why many researchers treat data as existing along a continuum rather than within a strict dichotomy, ranging from anecdotal or fictional to highly structured and statistical (Libarkin & Kurdziel, 2002). On this view, data are never purely quantitative, since both collection and interpretation are inevitably influenced by context and by the researcher's perspective. Also, qualitative evidence such as interviews or classroom observations can

be transformed into quantitative indicators through coding and content analysis, while survey data built on Likert scales are themselves shaped by subjective design choices.

The distinction between qualitative and quantitative, therefore, cannot rest solely on the type of data collected but must instead be understood in terms of the epistemological and ontological assumptions guiding inquiry, whether social phenomena are treated as measurable variables or as situated practices whose meanings require contextual interpretation.

These convergences indicate that the traditional divide between qualitative and quantitative research in SER is more porous than often assumed, opening the way for approaches that seek to combine them within a single design.

1.3 Mixed Methods research: canonical definitions and designs

As discussed in the previous sections, the history of SER has been largely shaped by the interplay between quantitative and qualitative traditions. Yet, the distinction between these orientations is more porous than often assumed, and researchers have at times combined elements of both within the same study (examples). From this perspective, mixed methods research occupies an important, if less explicitly acknowledged, place in the development of the field. Although qualitative and quantitative approaches have been more predominant, mixed-methods designs have also been employed in SER, reflecting the potential to capture different facets of science teaching and learning.

Their presence, however, has been comparatively weaker. A review of research trends in principal SER journals (Szyjka, 2012) noted the absence of references to mixed methods as a distinct categorization, suggesting that proposals for a "third paradigm" have not yet been consolidated as a universally accepted branch of empirical research. Instead, mixed methods often appear as pragmatic integrations of data sources, designed to enhance the robustness and implications of research outcomes.

Nevertheless, for the purposes of this thesis, mixed methods represent a crucial approach. Precisely because they challenge the strict opposition between qualitative and quantitative, they offer a methodological perspective from which to rethink how research in science education can be conducted, interpreted, and taught.

Looking at a more general panorama in social science, mixed methods research has emerged as an attempt to bridge the longstanding divide between quantitative and qualitative traditions. Since its early formulations (Johnson and Onwuegbuzie, 2004; Creswell, 2003), it has often been presented as a "third approach" designed to integrate the strengths of both paradigms. Grounded in a pragmatic orientation, mixed methods emphasize complementarity, combining statistical generalization with contextual depth, measurement with interpretation, numbers with narratives. This perspective resonates with the methodological history of SER, where the porousness of the qualitative–quantitative divide has created space for integrative designs.

Over the past two decades, the field of mixed methods research (MMR) has developed into a vast and consolidated area of inquiry. Today it has a dedicated journal (Journal of Mixed Methods Research), numerous methodological handbooks, and a wide range of manuals that have helped to establish its theoretical and practical foundations. Part of this trajectory can be traced to the early 2000s, when Creswell and Plano Clark contributed a chapter on "*advanced mixed methods designs*" to Tashakkori and Teddlie's Handbook of Mixed Methods in Social and Behavioral Research. As they later recalled,

"*It was our first endeavour to pull together our thoughts on mixed methods designs, and it was a good forum for presenting our ideas*" (Creswell et al., 2003). Their work was subsequently expanded and refined in later textbooks, which have become key references for both researchers and students.

The following sections explore the emergence of mixed methods research in education, tracing its philosophical foundations, its articulation as a paradigm, and its specific uptake in science education. Particular attention is given to both the promises and the limits of the pragmatic framing, which risks being insufficient, especially in today's research context, where entirely new forms of knowledge are being mixed. It is important to acknowledge that the field of mixed methods is vast and continually evolving; it would be almost impossible to map all the changes and refinements that have taken place over the past two decades.

The aim here is therefore modest: to highlight the basic ideas, to show where mixed methods started, and to examine the rationale that justified its use in education. This will provide a foundation for exploring where new forms of inquiry can be positioned and "named" with respect to mixed methods.

1.3.1 Mixed methods research in Education

In this section, I draw on *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research* (Creswell & Guetterman, 2019, Chapter 16), one of the most widely used manuals in educational research.

Mixed methods research is commonly defined as *"a procedure for collecting, analyzing, and 'mixing' both quantitative and qualitative methods in a single study or a series of studies to understand a research problem"* (Creswell & Plano Clark, 2018). Its basic assumption is that the combination of quantitative and qualitative methods provides a better understanding of the research problem than either approach alone (Creswell & Guetterman, 2019, p. 545). From this perspective, mixed methods require dual competence, as both traditions must be mastered, and the approach often entails substantial procedural demands (e.g., specialised skills, extensive data work, and, in some cases, teamwork).

A crucial element of this approach is integration. To make mixed methods research is not enough to collect and analyze two distinct strands. The "mixing" part is a key feature:

"It consists of merging (combining), connecting (having one database explain the other), building (having one database build something new to be used in the other), and embedding (placing one database within another larger database). In short, the data are "mixed" or "integrated" in a mixed methods study" (p.545).

Thus, the integration is conceived primarily as database-level work, where qualitative and quantitative strands are related through combination, sequencing, construction, or nesting.

Mixed methods study can be conducted in different situations, specifically *"when you have both quantitative and qualitative data, and these types of data, together, provide a better understanding of your research problem than either type by itself"*. That is because the combination of outcomes (quantitative) with processes (qualitative) is presented as yielding a "complex picture" of social phenomena. Also, a mixed methods design is indicated when a single approach is insufficient. For instance, when qualitative exploration is required to develop an instrument for subsequent quantitative testing, or when qualitative follow-up is needed to explain quantitative results. It is further suggested

that mixed methods offer alternative perspectives for applied contexts, such as policymaking, where both “numbers” and “stories” are demanded. Finally, reference is made to the increasing use of mixed methods in graduate training and in scholarly publications, sometimes as a pragmatic compromise in contexts where purely qualitative research is not yet widely accepted.

1.3.2 Basics mixed methods research designs

Different ways of combining qualitative and quantitative strands have been codified in the literature, resulting in a set of “core designs” that provide researchers with structured templates (Creswell & Plano Clark, 2018). Creswell and Guetterman (2019, p. 551) highlight three core models:

- The convergent design, in which qualitative and quantitative data are collected in parallel, analyzed separately, and then brought together for comparison, corroboration, or integration at the interpretation stage.
- The explanatory sequential design, in which the study begins with a quantitative phase, followed by a qualitative phase designed to explain or deepen the statistical results.
- The exploratory sequential design, in which the order is reversed: a qualitative phase is conducted first to explore a phenomenon, followed by a quantitative phase that tests or generalizes the initial insights.

Alongside these three foundational designs, they also describe a set of complex designs, which extend the logic of mixing into broader frameworks:

- The experimental design, which embeds qualitative strands into quantitative experiments to provide richer contextual understanding. In this configuration, the term *experimental* indicates that the quantitative component provides the primary structure of the study, typically involving a controlled intervention (pre/post testing, or comparison between groups). Whereas qualitative data are *embedded* to explore participants’ experiences, contextual factors, or mechanisms underlying the measured effects. This design is therefore particularly suited to studies seeking both causal evidence and interpretive insight.
- The social justice design, which encases a convergent, explanatory, or exploratory sequential basic designs with a framework (e.g., feminist or ethnic) to foreground marginalized voices and address issues of equity and transformation. Here, the adjective *social justice* refers to research explicitly oriented toward empowerment and emancipation. The mixed-methods configuration is not only instrumental but also value-driven, using methodological pluralism to highlight perspectives often excluded from dominant narratives. For example, participatory action research combining surveys and community interviews may serve to challenge inequities in STEM education
- The multistage evaluation design, which applies mixed methods across several phases of program evaluation, often involving cycles of qualitative exploration and quantitative measurement. The term *multistage* denotes the iterative structure of the design: data are collected and analysed in successive phases, with findings from one stage informing the next. The label *evaluation* underscores its applied purpose—assessing interventions, curricula, or policies over time to support refinement and evidence-based decision-making. This design exemplifies how mixed methods can sustain longitudinal inquiry and continuous improvement in educational contexts.

To make these models clearer and more immediately comparable, Table 1.1 summarises their key characteristics, while Table 1.2 presents examples of complex or extended designs and their typical research purposes.

Table 1.1 - Core mixed methods designs (adapted from Creswell & Plano Clark, 2018; Creswell & Guetterman, 2019)

Design type	Logic of timing	Priority between strands	Integration strategy	Typical purpose
Convergent design	Simultaneous collection and parallel analysis of quantitative and qualitative data	Usually equal priority	Merge or compare results at interpretation stage	To obtain a comprehensive understanding by validating or elaborating findings from both perspectives
Explanatory sequential design	Quantitative first, qualitative follows	Quantitative often prioritised	Connect by using qualitative phase to explain or expand on quantitative results	When quantitative results need contextual interpretation or elaboration
Exploratory sequential design	Qualitative first, quantitative follows	Qualitative often prioritised	Connect by using qualitative results to inform instrument development or hypotheses	When little is known about a phenomenon and qualitative exploration can guide later testing

Table 1.2 - Examples of complex or extended designs

Design type	Structure and logic	Distinctive feature	Example application
Experimental design	Qualitative component embedded within a quantitative experiment	Adds contextual depth and participant perspective to experimental results	Using interviews to interpret learning gains observed through tests
Social justice design	Classical design framed within a transformative paradigm (e.g. feminist, participatory, or decolonial)	Prioritises marginalised voices and social change objectives	Studying equity in STEM participation using surveys and participatory focus groups
Multistage evaluation design	Repeated cycles of qualitative exploration and quantitative assessment across project phases	Allows longitudinal evaluation and continuous improvement	Programme evaluation in education reform initiatives

This classification has played a central role in providing researchers with a structured repertoire for combining methods. It demonstrates that mixed methods can be organized into recognizable patterns, each with its own logic of timing, priority, and integration. At the same time, these designs reflect the pragmatic orientation of early mixed methods: the emphasis was placed less on resolving paradigmatic differences, and more on offering researchers practical templates for "what works" in combining qualitative and quantitative strands.

1.3.3 Mixed methods as "third paradigm"

As with qualitative and quantitative traditions, it is useful to return to the origins of mixed methods. A foundational contribution is the article by Johnson and Onwuegbuzie (2004), which explicitly introduced mixed methods as a "third paradigm" of research. Their contribution allows the researchers to understand that the emergence of mixed methods research is closely tied to the historical context of the "paradigm wars" of the 20th century.

During that period, a long-standing dispute between quantitative researchers, aligned with positivist or post-positivist assumptions, and qualitative researchers, aligned with constructivist or interpretivist assumptions, produced a climate of mutual distrust. Quantitative purists emphasized objectivity, causal determination, generalizability, and formalized writing styles. Qualitative purists, by contrast, defended multiple realities, context-bound knowledge, inductive reasoning, and empathic description.

Both camps often endorsed the *incompatibility* thesis, claiming that the two traditions could not, or should not, be combined within the same study. For quantitative purists, the concern was that introducing qualitative elements would undermine standards of objectivity, reliability, and generalizability, reducing scientific rigor to anecdote. For qualitative purists, the fear was the opposite: that bringing in quantitative measures would impose a reductionist lens, stripping away context, meaning, and the richness of lived experience. In both cases, the claim was anchored in deep epistemological commitments: to a single, measurable reality on one side, and to multiple, constructed realities on the other. The result was the development of two separate research cultures, which young researchers were often pressured to join.

Against this backdrop, Johnson and Onwuegbuzie (2004) argued that the division had become unproductive and that the overemphasis on differences had overshadowed areas of common ground. They proposed mixed methods research as a "third chair" alongside qualitative and quantitative research, envisioning as a way to combine the strengths and minimize the weaknesses of both traditions, able to offer a more realistic and useful account of how research is actually practiced. Their argument was that it should not be seen merely as a technical combination of qualitative and quantitative approaches, but as a distinctive paradigm, offering a unique way of engaging with complex research problems.

The rationale for this new paradigm was grounded in a call for methodological and epistemological pluralism. Contemporary research problems are complex, interdisciplinary, and socially embedded; they often demand flexibility that single-method approaches cannot provide. From this perspective, mixing methods offers advantages such as reducing bias, enabling corroboration, providing complementary insights, and broadening the scope of findings. Johnson and Onwuegbuzie further stressed that the link between method and paradigm is not fixed: quantitative tools can be used by qualitative researchers, and vice versa. A compatibilist, non-purist stance opens the way for designing studies that fit the research question rather than forcing the question into a pre-given mold.

This proposal was grounded in the philosophical stance of pragmatism, drawing on Peirce, James, and Dewey. Pragmatism treats the meaning and truth of ideas as inseparable from their practical consequences, offering researchers a way to move beyond metaphysical disputes about which paradigm is "right." In practice, this logic of inquiry can combine induction (discovering patterns), deduction (testing hypotheses), and abduction (inferring the best explanations for results). Researchers are encouraged to focus on the research question and consequently combine tools, techniques, and logics of inference as needed, without being constrained by the epistemological commitments of a single tradition.

However, this operation of mixing is delicate. It often involves the integration of knowledge traditions that originate in very different worldviews. This raises the question: on what assumptions, or vision of knowledge, can this mixing be legitimately based? The main answer offered in the early literature is the pragmatic vision articulated by Johnson and Onwuegbuzie. Yet, as Creswell and others have noted, this position has also been subject to criticism.

1.4 The Focus of this thesis

The previous sections have outlined the canonical traditions of educational research: first the qualitative and quantitative approaches, then the emergence of mixed methods as an attempt to bridge them. These accounts provide the necessary premises for positioning the focus of my research. My interest lies specifically in textual data analysis within SER, a field where qualitative traditions have always been central, but which is now encountering profound transformations through the rise of computational and AI-based methods. This intersection between established qualitative practices and new computational tools crystallizes the central methodological challenge that this thesis takes as its vantage point.
At the same time, because this discussion unfolds within science education research, my focus must also be situated within the broader paradigmatic landscape of the field. As Treagust and Won (2023) have argued, SER cannot be understood only in terms of methods and techniques: it must be interpreted through the paradigms that shape its foundational assumptions, which in the case of science education are positivist, interpretivist, critical, and pragmatic. In this sense, my analysis is aligned with their approach: it is not limited to describing methods but seeks to uncover and make explicit the epistemological commitments that lie beneath methodological choices. This orientation ensures that the exploration of textual data and computational methods is not treated as a purely technical innovation, but as part of a deeper paradigmatic reflection on the recent and future of research in science education.

1.4.1 Textual data, and computational AI-based methods for SER

The focus of this thesis on textual data in science education research was prompted primarily by my own research experience, which I believe is shared by many other researchers. During my master's studies and later as an early-stage researcher within a founded project, I worked extensively with textual data such as teachers' interviews and students' essays. In both cases, the datasets were of medium size, consisting of hundreds of minutes of interviews and hundreds of pages of text. My expertise at that stage was grounded in qualitative methods, particularly thematic analysis, and between 2020 and 2021 I encountered what many qualitative researchers experience: the *struggle* of analyzing rich, nuanced data that require time, care, and constant interpretative decisions. That struggle led to valuable insights, but at the same time, with my background in physics and an enduring

interest in computation, I became increasingly aware that new computational tools were emerging that could be applied precisely to this kind of data.

A second motivation comes from a more personal methodological curiosity. Beyond my empirical work, I have always been drawn to the beauty of methodology itself, to the questions and challenges that arise when different approaches to analysis are employed. This curiosity led me to connect the dots between my qualitative expertise, the rise of computational/AI methods, and the specific methodological challenges posed by textual data in science education. Together, these elements constitute the central focus of this thesis.

Features of textual data in educational context

In SER, textual data are among the most widely used forms of evidence. One simple reason is that the voices of people (such as students, teachers, and other stakeholders) are crucial for collecting information about the experience of science learning, students' difficulties, their ideas, and the needs and beliefs of the educational community. This focus intensified as SER shifted its attention toward situated contexts (see Section 1.2.2), when the priority moved from theory verification and generalized outcomes to the understanding of phenomena within their specific and contextualized settings (Szyjka, 2012). As this interest grew, qualitative text-based data became a fundamental resource for educational inquiry, providing critical access to how learning processes unfold, how they are perceived by participants, and how they are shaped by local contexts (Caramaschi et al., 2025).

Focusing on the specific forms of qualitative textual data that are routinely collected and analyzed in SER, we find, for example, oral interviews on certain topics conducted by researchers and subsequently transcribed, written comments gathered alongside quantitative data through questionnaires, and students' written tasks or essays on scientific topics. In addition, open responses provided by students in school tests and the textual content of textbooks or other instructional materials supplied by teachers can also serve as valuable data sources. In practice, this means that textual data are embedded in the everyday work of researchers in science education.

At this point, it is important to highlight two characteristic traits of textual data in SER when they are employed for research purposes:

- Volume of data. The amount of available material in a single study can be moderately large-for example, in terms of the number of texts or their length. While such datasets do not qualify as *big data*, they are nevertheless substantial enough to make traditional qualitative analysis time-consuming.
- Context-specificity. Textual data are usually collected within a particular context or tied to a specific task. As a result, texts often display topics explicitly or implicitly connected to that context. Sentences tend to be moderately articulated in terms of semantics, and domain-specific language is frequently used.

These features, combined with the inherent complexity of language, pose significant challenges for researchers working with textual data in SER (Caramaschi et al., 2025). For instance, Stanley and Robertson (2024) reviewed current qualitative SER literature with a focus on methodological aspects, showing that case studies often generate high volumes of textual data, sometimes reaching dozens of hours of audio recordings and hundreds of minutes of interviews. Another recurring methodological challenge arises from the richness and specificity of language itself: decisions such as how to reduce the dimensionality of the text (Mariegaard et al., 2022) or which coding strategy to apply (Wulff et al.,

2023) become highly consequential, as they directly affect the interpretability, reliability, and reproducibility of results.

The possibility of AI and computational methods

As described above, unstructured textual data in SER have typically been analysed using qualitative methods. However, the increasing digitalization of education and the rapid development of data science techniques are opening new possibilities. In this context, computational, data-intensive methods emerge as an appealing alternative for exploring the large volumes of textual data now available in SER.

These methods include the use of machine learning algorithms to classify or retrieve portions of a corpus under study, as well as a wide range of natural language processing and statistical approaches to language. Such tools promise not only to increase the efficiency of analysis but also to reveal patterns and structures in textual data that are difficult to access through traditional qualitative approaches alone.

1.4.2 Paradigms in Science Education Research

The analysis of textual data and the emergence of computational, AI methods in science education research cannot be understood solely as technical innovations. As Treagust and Won (2023) remind us, research in SER must be situated within paradigmatic frameworks that define how researchers conceive of reality, knowledge, and method. Paradigms are not simply sets of tools; they are comprehensive belief systems that carry ontological, epistemological, and methodological assumptions. Asking what reality is, whether and to what extent it is knowable, and how it can be investigated are not abstract philosophical games-they are questions that directly shape the ways in which SER is conducted.
Treagust and Won identify four major paradigms currently shaping science education research: positivist/post-positivist, interpretivist, critical theory, and mixed methods.

Positivism and post-positivism regard reality as existing independently of the researcher. In its classical form, positivism assumes a knowable, law-like social reality (naive realism), whereas post-positivism (or critical realism) recognizes the imperfect and probabilistic nature of our knowledge. Research within this paradigm emphasizes explanatory designs such as experiments or surveys, large-scale assessments, quantitative measurement, and validity checks. Researchers are positioned as objective, distant collectors of data, and quality is assessed through reliability and validity.
Interpretivism/constructivism, by contrast, denies the existence of an objective, independent social reality. Reality is seen as constructed through meaning-making processes and multiple interpretations (constructionism, relativism). Here, research emphasizes exploratory designs such as ethnography, phenomenology, or grounded theory, conducted in naturalistic settings. The researcher is not detached, but an insightful interpreter and storyteller, co-constructing meaning with participants. Quality is established through thick description, prolonged engagement, reflexivity, and checks such as triangulation or member review.

As outlined in Section 1.2.2, these two paradigms have historically been placed in opposition, producing the long-standing debate between quantitative and qualitative traditions. Yet SER has also been shaped by two additional paradigms that complicate this dichotomy.

Critical theory emphasizes the role of research in uncovering power relations, questioning taken-for-granted assumptions, and supporting the transformation of educational structures for marginalized groups. It seeks not only to describe learning, but to emancipate the disempowered, often through action research in historically disadvantaged contexts. This paradigm foregrounds the ethical and political dimensions of SER and brings questions of equity, participation, and justice to the center of methodological design.

Mixed methods are identified by Treagust and Won as the fourth paradigm, albeit with some caution. Unlike the others, it is not underpinned by a single coherent philosophical worldview but is instead often conflated with research designs or methods. Nevertheless, they argue that mixed methods can be seen as a paradigm in its own right, insofar as it represents a different way of framing research. By combining quantitative and qualitative data, mixed methods approaches claim to transcend the limitations of either tradition alone. Yet this raises critical questions: Can realism, relativism, and critical perspectives truly be integrated within one study? Can a researcher be at once a distant observer, a passionate interpreter, and a social advocate? These provocative questions signal the need to avoid treating mixed methods as a purely technical combination of tools and instead recognize its deeper philosophical implications.

For this thesis, these paradigms matter because they remind us that textual data analysis, whether qualitative or computational, is never neutral. Choices about whether to rely on coding strategies, statistical language models, or machine learning classifiers cannot be separated from assumptions about what counts as knowledge, what language represents, and what kind of generalizations are sought. Following Treagust and Won, I adopt an approach that consistently seeks to make explicit the assumptions that underlie methods of analysis. In this sense, my investigation of textual and AI-based methods is not merely about comparing techniques, but about situating them within the paradigmatic debates that continue to shape science education research.

1.5 Transformations in the era of data science and AI

Up to this point, the discussion has hinted at (but not yet explicitly addressed) the profound transformations currently underway that have brought methodological debates back to the forefront. These changes invite us to reconsider research methods not merely at a technical level, but at a foundational one. In this section, I provide an overview of the broader transformations reshaping society, and consequently SER.

1.5.1 Changes impacting the way research is done in the field of Science and Physics Education

Over the last decades, the field of SER has been affected by an epochal shift toward data collection and analysis that has influenced nearly all research domains across society. These changes have been driven by several key phenomena, first and foremost the “datafication” of socio-institutional contexts, such as education, which has considerably expanded the variety of available digitised data (Jarke, & Breiter, 2019).

In educational contexts, digital platforms, online environments, and technological tools have been increasingly incorporated into traditional settings and practices, contributing to a shift toward digitalisation and data-driven systems. Concurrently, we have witnessed the so-called *Big Data* revolution: data has become available in unprecedented amounts and variety, and technological

infrastructures and tools now allow their rapid collection and storage in digital format (Klašnja-Milicevic et al., 2017).

This transformation coincides with what Wise and Shaffer (2015) describe as the *Big Data* era in education, an era in which the increasing digitisation of learning environments provides vast and fine-grained datasets. These online learning environments generate rich traces of learner behaviour, interactions, and language. Technological developments now facilitate the collection of extensive data at various levels (individual, classroom, and national) about potentially all processes of teaching, learning, and school management (Jarke & Breiter, 2019). Fischer and colleagues (2020) identified three main big data typologies coming from educational contexts:

a. The *microlevel* (e.g., clickstream data), which includes "*fine-grained interaction data with seconds between actions that can capture individual data from potentially millions of learners*".
b. The *mesolevel* (e.g., text data), which includes "*computerized student writing artefacts systematically collected during writing activities in a variety of learning environments ranging from course assignments to online discussion forum participation*".
c. The *macrolevel* (e.g., institutional data), which includes "*student demographic and admission data, campus services data, schedules of classes and course enrollment data*".

As Romero and Ventura (2020) note, the scale and diversity of these data have necessitated increasingly automated analytical approaches. The growing use of data science techniques in various socio-economic fields, including education, has made data-intensive methods a compelling solution for processing large datasets computationally and rapidly (Fischer et al., 2020). The so-called "big data" revolution has thus fuelled the scientific and technological development of new analytical approaches such as machine learning and natural language processing (Fischer et al., 2020).

In recent years, this situation has led to the emergence of new interdisciplinary communities operating at the intersection of data science and education: Educational Data Mining (EDM) and Learning Analytics (LA) (Baker & Siemens, 2014). These communities arose in the mid-2000s as responses to the methodological challenges posed by large-scale educational data. The first *Educational Data Mining* conference was held in 2008, followed by the inaugural Learning Analytics and Knowledge (LAK) meeting in 2010, which later led to the creation of the *Society for Learning Analytics Research (SoLAR)* (Siemens & Baker, 2012). Both communities share the goal of improving education through data-driven insight, though they differ in orientation: LA emphasises human interpretation and pedagogical decision-making, whereas EDM tends toward automated discovery and reductionist modelling (Baker & Inventado, 2014). These developments have supported the evolution of a broader research area known as Education Data Science, where statistical and computational techniques are employed to explore educational questions and phenomena (McFarland et al., 2021). Specifically for textual data, the field of Educational Text Mining has been established as a domain for analysing large-scale textual data within educational settings.

Together, these transformations have opened up vast possibilities for rethinking how research is conducted within Science Education Research. Data science techniques are now being adopted for data analysis and inquiry, expanding the methodological repertoire available to researchers. However, theoretical contributions on the methodological implications of these practices remain underdeveloped. The present thesis therefore aims to contribute to deepening methodological reflection and advancing the study of methods in SER.

1.5.2 Different approaches for textual data analysis in Science Education Research (SER)

The tradition: qualitative approaches

Traditionally, textual data in SER have been analysed using qualitative approaches. Although not always explicitly stated in articles, qualitative approaches have been historically generated within a specific paradigm, the interpretivist one (Treagust & Won, 2023). A paradigm is based on ontological, epistemological, and axiological assumptions, about the reality under study, its knowability, the type of research problems, the values and philosophical influences, the research design and methods, and the role and practice of researchers. The paradigm plays a crucial role because it "*guide[s] the researchers throughout the empirical research process: from setting the research purpose to selecting data-collection methods to analyzing the data to reporting the findings*" (Treagust & Won, 2023, p.3). In the opening chapter of a recent methodological handbook, Treagust and Won argue that the interpretivist paradigm stems from relativist ontology and constructivist epistemology (Treagust & Won, 2023), where reality is not considered objective per se, but knowable through the meaning attributed by individuals, admitting different interpretations. The research within this paradigm foresees an exploratory design. The usual research topics addressed with this approach are lived experiences of teachers and students. The quality of the interpretation process is ensured by a prolonged meaningful interaction in the context and repeated re-analysis and reflections on data. Additional elements that guarantee the process validity are the researcher's skills, sensitivity, and integrity, member checking, audit trail, thick descriptions, peer debriefing, triangulation, and negative case analysis (Anfara et al., 2002).

The computational data-driven approaches

The digital shift and technological revolution and in turn, the wide rise of Education Data Science have contributed to open new possibilities for textual data analysis in SER (Fesler et al., 2019). Alongside a more traditional qualitative paradigm, an alternative analytical approach emerged. The computational data-driven approach "*is supposed to be data-driven, strongly inductive, and relatively theory-independent*", as well as largely automated (Pietsch, 2015). It assumes, more or less implicitly, the idea that data in big amounts is a necessary and sufficient container of knowledge and that implicit patterns in the data can be highlighted only by the analysis of the data itself. Thus, a specific theoretical design at the basis of data analysis is no longer needed to orient the extraction of valuable information from data.

Belonging to this approach, we refer to all those methods falling under Artificial Intelligence (AI) techniques for textual data analysis, such as Natural Language Processing (NLP) techniques, Machine Learning (ML), and Network Analysis which are precisely those "*new and often non-traditional*" methods included under the umbrella of Education Data Science (McFarland et al., 2021, p.2). In social sciences research, the approach consists of data collection and pre-processing, followed by data analysis, in which traditional statistics methods and the most widely employed ML methods are usually combined. Finally, the validation phase is needed to ensure correctness and accuracy (Zhang, et al., 2020).

In education research, articles containing keywords such as ML, AI, Learning Analytics, Data Science, NLP, and Network Analysis have massively grown since 2010 (McFarland et al., 2021). Even in SER, interesting works adopted this kind of approach for dealing with textual data (e.g. Wulff, et al., 2023; Odden, et al., 2021; Bruun, et al., 2019). Although strengths started to be highlighted, there are also criticisms and increased awareness of the intrinsic limitations that characterize this approach (Pietsch, 2015). Therefore, in SER, where the debate is still at its

beginning, the deepening of the paradigm behind these methods (i.e. ontological, epistemological, and methodological assumptions) is felt as an urge.

1.6 Impacts on the paradigms

1.6.1 Calls for "a critical examination" of quantitative and qualitative methods

In the current data-intensive and socio-technical research era, the conscious use of methods in SER requires a deep understanding of their benefits, limitations, and conditions of applicability. In recent years, a growing interest has emerged in examining how methods are articulated within the field.

A particularly significant example comes from Physics Education Research (PER), where the journal *Physical Review Physics Education Research* launched two focused collections explicitly dedicated to methodological reflection. The first, published in 2017, was entitled *Quantitative Methods in PER: A Critical Examination*, while the second, in 2021, addressed *Qualitative Methods in PER: A Critical Examination*. The very choice of subtitle signals a turning point: both traditions were to be examined critically rather than simply employed as taken-for-granted frameworks.

These initiatives mirror a broader concern within the SER community to re-assess both paradigms. On the quantitative side, the expansion of available data sources and analytical techniques has raised the need to reconsider how quantitative methods are applied, with Henderson (2017) emphasising the importance of supporting researchers in making more informed methodological choices. On the qualitative side, decades of development and diversification, from classical ethnographies to newer interpretive and constructivist approaches, have made it timely to provide a mapping of strengths and weaknesses in this area as well (Henderson, 2021).

Taken together, these contributions highlight several elements of novelty currently reshaping the field. One concerns new contexts: quantitative research in PER has expanded beyond traditional measures of achievement and demographics to include *"even behavior both online and in the classroom"* (Henderson, 2017). Such settings generate new forms of data and open questions about how learning and teaching should be studied in hybrid or digital environments. Qualitative studies, likewise, are increasingly situated in diverse contexts that combine in-person and online spaces. These environments offer opportunities to capture the lived experiences of learners but also demand critical reflection on how researchers engage with participants and how context shapes interpretation.

Another element of novelty relates to methods and techniques. The editors of the quantitative collection observed that "*many tools and analysis techniques have become standard ways to measure student learning and outcomes, faculty teaching behaviors, departmental changes, and more. As PER matures and research progresses, new tools and analysis techniques are becoming available*" (Henderson, 2017). This expansion brings with it the challenge of determining which tools are appropriate for which studies, and how differences among them affect the data produced and the conclusions drawn. On the qualitative side, alongside traditional sources such as interviews and focus groups, researchers now rely on portfolios, digital transcripts, and other artefacts that broaden the repertoire of qualitative inquiry. New lenses and analytical procedures, including computational approaches to coding and interpretation, further diversify the field and raise pressing questions about validity, reliability, and the comparative affordances of different strategies.

These developments ultimately reshape the kinds of research questions that can be asked, the ways evidence is collected and analysed, and the assumptions researchers make about what can be known in educational contexts.

1.6.2 Examples that break the dichotomy between qualitative and quantitative

The theoretical impasse outlined above is made evident by research practices that challenge the validity of traditional paradigmatic distinctions in SER. Building on earlier reflections on the distinction between post-positivist and interpretivist paradigms (Fantini, 2014), a series of examples have been developed, to illustrate how contemporary practices blur the dichotomy. In particular, it is shown how new techniques, forms of data, and settings question the assumptions that underpinned both quantitative and qualitative (Caramaschi, et al., 2023).

Example A - Data Mining and the "loss of control" in quantitative research

A first case comes from Ding's (2019) taxonomy of quantitative methods, which lists three genres of quantification: (a) *measurement*, or the quantification of individual constructs; (b) *controlled exploration of relations*, or the quantification of relationships between constructs through designed studies; and (c) *data mining* (DM), or the quantification of new information from large datasets. Ding (2019) explicitly characterizes data mining as "*the newest line of work*" among quantitative genres, signalling its novelty. This genre refers to computational techniques increasingly used in educational research, particularly in EDM and Learning Analytics (Zorić, 2020; Romero & Ventura, 2013, 2020).

It is precisely this third genre that reveals a paradigmatic tension. Measurement and experimentation are typically theory-driven: variables are defined in advance, data are deliberately collected, and hypotheses are tested under controlled conditions. DM, by contrast "*differs from the previous in that investigators in this tradition often have little control of data collection. In fact, the data being examined often have been previously collected*" (Ding, 2019, p. 8). This highlights that DM often operates on pre-existing datasets that were collected without the explicit supervision or a priori choices of the researcher (e.g., log files from online platforms, demographic databases, institutional records).

This "loss of control" challenges one of the axioms of quantitative research, namely that variables and conditions are designed by the investigator to enable hypothesis testing. Instead, DM proceeds in a more exploratory manner, with patterns emerging inductively from aggregation, closer to interpretivist traditions. Ontologically, this assumes that human behaviour manifests itself in patterns only visible through large-scale aggregation. Epistemologically, it displaces theory-driven variable selection with researcher choices during data pre-processing and algorithmic analysis.

The inclusion of DM under the quantitative umbrella therefore destabilises the very boundary it is supposed to represent: a method labelled "quantitative" rests on assumptions that blur toward interpretivism. As Ding's taxonomy shows, the quantitative tradition has adapted to the unavoidable entry of big data into SER, but without fully re-examining its paradigmatic bases.

Example B - Log Data and the challenge to the "natural setting"

A second case concerns the increasing use of log data, digital traces automatically recorded when learners interact with educational platforms. These traces can capture, with unprecedented detail, how long students spend on tasks, where they direct their attention on screen, or which resources they

consult most frequently (Wang et al., 2023). These data are multiple and varied (e.g., clickstream, eye-tracking, online forum posts, demographic and institutional records (see details about these data in section 1.5.3). By combining behavioural traces with surveys or psychological scales, researchers can map action sequences to cognitive traits and refine theories at a granular level.

However, these practices destabilise a foundational assumption of qualitative research: the notion of the *natural setting*. Approaches such as grounded theory presuppose immersion in context, where data collection and analysis are progressively refined, driven by the awareness of the contextual elements and/or by the experience of direct engagement with participants. Log data, by contrast, are generated independently of the researcher's presence, often in hybrid online–offline environments where the boundaries of context are blurred and difficult to delimit. As Fischer et al. (2020, p. 143) caution, "*Researchers cannot ignore contextual factors such as the stimuli to which students are responding. If researchers do not pay attention to unique contextual factors, techniques for analyzing meso-level big data might result in inaccurate inferences*".

Ontologically, this undermines the assumption of a stable, "natural" setting for qualitative inquiry. Epistemologically, it complicates interpretation: the richness of human experience is mediated by digital traces that are simultaneously behavioural, contextual, and decontextualised.

Taken together, these two examples show how computational practices and digital settings destabilise both poles of the classical dichotomy. Data mining troubles the post-positivist logic of quantitative research by introducing exploratory, theory-weak analysis of pre-existing datasets. Log data challenge the interpretivist foundation of qualitative research by undermining the assumption of the natural setting. This shift reveals how novel computational practices and digital environments no longer fit the ontological and epistemological bases of traditional research, questioning the applicability of their classical assumptions in the current era.

1.6.3 Blurring boundaries and re-thinking paradigms

The cases discussed above illustrate examples of new conditions and contexts that do not fully align with the assumptions of traditional approaches. These include, for instance, the need to extend the epistemology of the educational context to account for transformations in the very nature of data, in the modalities of data collection, and in the role of the researcher.

This situation invites a re-thinking of methodological paradigms in SER, opening epistemological and ontological questions that can stimulate reflection. For instance: *Is it possible to add an interpretative layer in data mining? If so, how, and where? How can the basic assumptions of a grounded theory approach be redefined in order to consider also digital platform data?*

These questions exemplify the need to revisit the assumptions that sustain our methodological choices. We highlight that both examples seem to express the idea that qualitative elements are more and more integrated with quantitative ones. Thus, we raise the issue of whether mixed methods research can provide a framework for addressing such blurred cases, not merely as a technical juxtaposition of procedures, but as an approach capable of articulating situations where elements of one tradition encounter the other.

The next section therefore turns to the concepts of mixed methods research in education, offering an overview of how this approach has been theorised and under what conditions it can integrate elements of qualitative and quantitative inquiry within the field of science education.

1.6.4 Expansions and changes within mixed methods research

Recent developments within the mixed methods community reveal how the field itself is expanding to accommodate new methodological frontiers. These transformations are not only driven by technological innovation but also by a rethinking of the paradigmatic foundations of "mixing".

A striking example comes from the *Journal of Mixed Methods Research*, the flagship journal of the discipline, which issued in November 2024 a Call for Papers for a special issue entitled *"Advancing Qualitatively Driven Mixed Methods Research: Tackling Societal Challenges"*.

The editor describes this as *"a pivotal moment in the history of mixed methods research inquiry",* recognising that for a long time the field has largely *"relied solely on quantitatively driven mixed methods inquiry rooted in positivist and post-positivist methodologies emphasizing hypothesis testing, causality, observation, and measurement". It is observed that over time this dominant orientation has "limited our angle of vision onto the social reality to only those issues we can understand through observation and precise measurement"*.

The call strongly reaffirms the importance of qualitatively driven mixed methods, recognising the unique contribution of qualitative approaches in revealing unheard voices and plural perspectives of lived experience. This point is particularly relevant to the present thesis, which acknowledges that within qualitatively driven designs, textual and narrative forms of data play a crucial role as principal sources of information. As the guest editors explain, the field has *"historically ignored the importance of qualitatively driven approaches to mixed methods inquiry that offer a unique lens through which we can explore lived experiences by revealing subjugated knowledge that quantitatively driven mixed methods inquiry fails to capture"*. From this perspective, qualitatively driven inquiry provides a deeper understanding of complex societal challenges (e.g. inequality, marginalisation, and health disparities) whose very focus lies in the analysis of lived experience as a critical dimension of social problems.

Importantly, the call also highlights the integration of AI into qualitatively driven inquiry, positioning it as a key methodological frontier. In fact, the community is invited to contribute by *"exploring qualitatively driven methodological, theoretical, and research design innovations"* and by *"leveraging and integrating new digital technologies like artificial intelligence to enhance qualitatively driven inquiry"*.

The fact that the leading journal in the field places the combination of AI and qualitatively driven designs at the centre of its agenda demonstrates that this is no longer a marginal conversation but one formally recognised by the institutions shaping mixed methods scholarship. This evolution mirrors the broader transformations occurring across the social sciences, where interpretive and computational epistemologies are increasingly interwoven.

Another significant signal of change is found in *The Routledge Reviewer's Guide to Mixed Methods Analysis* (edited by Onwuegbuzie and Johnson, 2021) which introduces innovative ways of conceptualising mixed methods. Among these are the notions of crossover mixed methods and inherently mixed methods, both of which expand the classical Creswell-Plano Clark typology. The first refers to designs that intentionally integrate procedures or analytical logics typically associated

with one strand into the other. For instance, quantitative coding is done using qualitative data. This reflects the growing permeability between methodological boundaries.

In Chapter 1 the same book, a series of mixed analysis formulae is described to illustrate how qualitative and quantitative approaches can be combined to analyse either qualitative or quantitative data (Onwuegbuzie and Johnson, 2021). One of these, the crossover mixed analysis, extends traditional conceptions of mixing by proposing that *"one or more analysis types associated with one tradition (e.g., qualitative analysis) are used to analyse data associated with a different tradition (e.g., quantitative data)"* (Onwuegbuzie & Combs, 2010). Because these formulae are technically detailed, readers are referred to the original source (see Chapter 1 of Onwuegbuzie & Johnson, 2021, p. 2) for a full account.

The second notion, inherently mixed methods, describes approaches in which mixing is embedded at every stage (conceptual, analytical, and interpretive), making any separation between strands impossible. This view implies a paradigmatic rather than merely procedural understanding of mixing. In line with Bazeley (2018) and her related writings on integration in mixed methods, *inherently mixed* methods are those whose analytic logic already fuses qualitative and quantitative elements within a single procedure. Here, integration is built into the method itself: meaning-making (qualitative) and pattern or measurement (quantitative) operate together rather than in separate stages. In this sense, the method is *"born mixed",* and the qualitative-quantitative distinction is inseparable within one analytic process.

For clarity, this term differs from another type of mixed methods called hybrid methods, which combine distinct analytic procedures by design, typically one qualitative and one quantitative, with the researcher orchestrating how they inform one another. Integration thus occurs *across* methods, not *within* a single inseparable procedure. The techniques remain identifiable (for instance, a qualitative analysis paired with a statistical or algorithmic analysis), but they are deliberately linked so that their findings become interdependent for interpretation.

A quick test to distinguish these two types of mixed analyses is summarised in Table 1.3, adapted from Bazeley (2018), to highlight *the location and mechanism of integration* in the inherently and hybrid methods.

Together, these developments indicate that the integration of computational methods and qualitative methods and the division between qualitative and quantitative boundary is no longer speculative but a recognised frontier within the discipline.

Table 1.3 - Quick test to tell inherently mixed and hybrid methods apart

Question	Inherently mixed	Hybrid
Where does integration live?	Inside one method's core logic.	Across multiple methods you combine.
Are the techniques separable?	No. Co-constitutive within a single procedure.	Yes. distinct analyses, linked by design.

1.7 Rethinking mixed methods today

The classical conception of mixed methods research was developed at a time when the main challenge was how to combine two distinct methodological traditions: quantitative and qualitative. As outlined in Creswell and Plano Clark's designs, integration was typically framed as a matter of sequencing or prioritizing two datasets, for example, collecting survey data followed by interviews, or merging parallel quantitative and qualitative analyses. The central questions were largely technical: "Which method should come first?", "Which type of data should be given priority?", "How can results from different strands be meaningfully combined?".

While this framework has been influential, it becomes insufficient in the context of today's computational and AI-based approaches. Integration is no longer just about juxtaposing two separate datasets. Instead, new practices often involve reconfiguring a single dataset through multiple methodological lenses. A paradigmatic example is textual analysis: a corpus of student writing or interview transcripts may be examined qualitatively through thematic coding, but the same texts can also be transformed into numerical vectors using word embeddings or other natural language processing techniques. In such cases, there is no strict "sequence" of separate data collections. Rather, the same dataset is recast into different representational forms, each carrying its own assumptions about what language is and how it can yield knowledge.

This shift points to a deeper epistemological novelty. Classical mixed methods research tended to foreground the type of data, quantitative or qualitative, as the defining element of methodological choice. But in computational/AI hybrids, the key issue is not simply the type of data, but the way in which data are represented, processed, and interpreted. Language, for instance, is treated simultaneously as measurable (a sequence of numbers or vectors) and interpretable (a carrier of situated meaning). Such a dataset-centric hybridity cannot be fully captured by designs that assume two independent strands of inquiry.

Recognizing this novelty also forces a reconsideration of the ontological and epistemological assumptions embedded in research. Computational approaches are not simply "new tools"; they reshape how knowledge is produced. They raise questions about what counts as interpretation, what level of generalization is possible, and how human meaning-making is related to algorithmic pattern-finding. In this sense, hybridity in methods also entails hybridity in the very logics of inquiry: deduction, induction, and abduction may all be involved, but in unfamiliar combinations.

For these reasons, it is no longer adequate to treat mixed methods as a neutral technical toolkit for combining procedures. Instead, they must be understood as value-laden and knowledge-producing practices that embed specific ontological, epistemological, and even axiological commitments. Novel approaches to mixed methods, including those that emphasize crossover or inherent hybridity, gesture in this direction, but the debate is still in its early stages. What is required now is a more paradigmatic reflection — one that addresses the contexts in which these new methods can be developed, the kinds of data and data transformations they presuppose, the assumptions introduced by computational approaches that are not reducible to classical statistics, and the conditions under which computational and qualitative insights can be meaningfully integrated.

The critical task, then, is not only to refine methods but to reconsider the very frameworks within which "mixing" is understood. This thesis takes up that challenge in the specific domain of science education research, focusing on the methodological and epistemological implications of combining computational and qualitative approaches to textual data.

1.8 Conclusions

In this chapter, I have sought to provide a broad methodological map of the approaches and research paradigms that shape science education research. This overview was necessary in order to situate the focus of the thesis: the emergence of new approaches to textual data analysis that challenge the traditional boundary between qualitative and quantitative inquiry. Within the mixed methods tradition, new forms of analysis have already been proposed that resemble the kinds of innovations I intend to investigate in the field of SER.

At the same time, today's research landscape is being reshaped by profound societal and technological changes, from the digitalization of education to the rise of artificial intelligence. These shifts introduce novel elements, both contextual and methodological, that demand careful reflection. They call for a perspective, such as that of Treagust and Won (2023), which insists on ontological, epistemological, and methodological awareness as the basis for any responsible methodological innovation.

From this perspective, a number of pressing questions arise. What kind of approach is constituted when computational and qualitative methods are combined in SER? What paradigmatic framework, if any, can adequately embrace such an approach? Can mixed methods research serve this role, and if so, what kind of mixed method is it, and what assumptions does it embed? How should we justify the mixing of computational and qualitative methods? And finally, are computational approaches to language best understood as extensions of quantitative methods, or do they represent something different altogether?

These questions set the agenda for the following chapters. Each of the next four chapters addresses one dimension of this problem, exploring how computational and qualitative approaches can be combined in science education research, and how such combinations might contribute not only to methodological innovation, but also to a rethinking of the mixed methods paradigm itself.

There is a *methods map* on SAGE Research Methods website (Figure 1.2).
It says: "*Mixed methods. Inquiry that combines two or more methods*".

Today, in science education research, we have precisely that: qualitative approaches on one side, and computational/AI-based approaches on the other.

The same source continues: "*This particular term usually refers to mixing that crosses the quantitative-qualitative boundary*".

I believe there is no better description of what is currently happening in SER when machine learning, network analysis, or natural language processing are combined with qualitative research:

We are crossing boundaries.

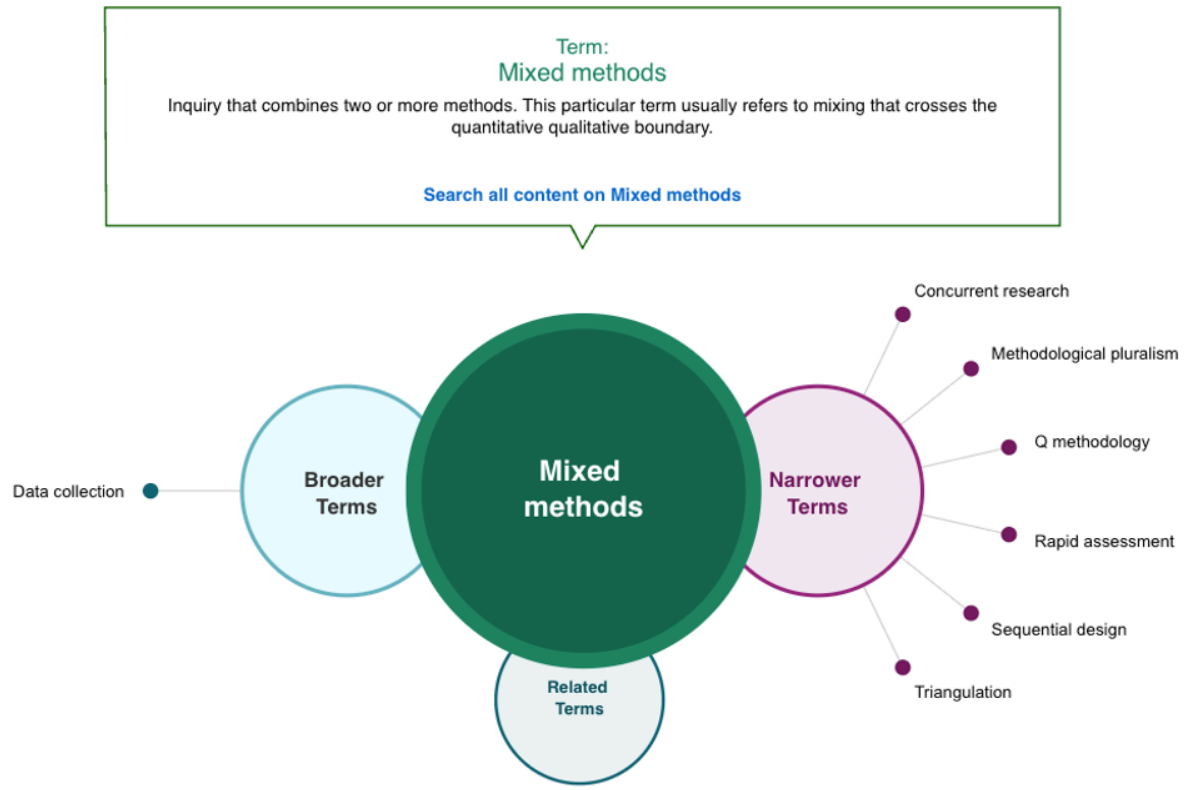

Figure 1.2 - *Methods map* on SAGE Research Methods website. Source

References

Anfara, V. A., Brown, K. M., & Mangione, T. L. (2002). Qualitative Analysis on Stage: Making the Research Process More Public. *Educational Researcher*, 31(7), 28-38. https://doi.org/10.3102/0013189X031007028

Baker, R., & Inventado, P. (2014). Chapter 4: Educational Data Mining and Learning Analytics. In J.A. Larusson and B. White (eds.), Learning Analytics: From Research to Practice, DOI 10.1007/978-1-4614-3305-7_4, © Springer Science+Business Media New York 2014

Baker, R., & Siemens, G. (2014). Educational Data Mining and Learning Analytics. *In: Sawyer RK, editor. The Cambridge Handbook of the Learning Sciences.* Cambridge: Cambridge University Press; p. 253–72. (Cambridge Handbooks in Psychology) https://doi.org/10.1017/CBO9781139519526.016

Bazeley, P. (2018) Integrating Analyses in Mixed Methods Research. SAGE Publications Ltd., Thousand Oaks. https://doi.org/10.4135/9781526417190

Bogdan, R. C., & Biklen, S. K. (2003). Qualitative Research of Education: An Introductive to Theories and Methods (4th ed.). Boston: Allyn and Bacon.

Bruun, J., Lindahl, M., & Linder, C. (2019). Network analysis and qualitative discourse analysis of a classroom group discussion. *International Journal of Research and Method in Education*, *42*(3), 317-339. https://doi.org/10.1080/1743727X.2018.1496414

Call for Papers for a Special Issue on “Advancing Qualitatively Driven Mixed Methods Research: Tackling Societal Challenges.” (2024). Journal of Mixed Methods Research. https://doi.org/10.1177/15586898241302983

Caramaschi, M., Fantini, P., & Levrini, O. (2023). Do the epistemological bases of qualitative and quantitative paradigms in STEM education still hold today? Oral contribution. *Measurement in STEM Education* conference. https://indico.unina.it/event/60/contributions/638/

Caramaschi, M., Zanellati, A., Levrini, O. (2025). Comparing Textual Data Analysis Methods in Science Education Research. Chapter in “Connecting Science Education with Cultural Heritage” Springer book. DOI: 10.1007/978-3-031-90944-3

Combs, J. & Onwuegbuzie, A. (2010). Describing and Illustrating Data Analysis in Mixed Research. International Journal of Education. 2. 10.5296/ije.v2i2.526.

Creswell, J. W. (2003) Research Design: Qualitative, Quantitative, and Mixed Method Approaches. Sage Publications, Thousand Oaks.

Creswell, J. W., & Guetterman, T. C. (2019). Educational research: Planning, conducting, and evaluating quantitative and qualitative research (Sixth edition). Pearson.

Creswell, J. W., & Plano Clark, V. L. (2018). Designing and Conducting Mixed Methods Research (3rd ed.). Thousand Oaks, CA: SAGE.

Denzin, N. K., & Lincoln, Y. S. (2000). Introduction: The Discipline and Practice of Qualitative Research. In N. K. Denzin, & Y. S. Lincoln (Eds.), Handbook of Qualitative Research (2nd ed., pp. 1-34). Thousand Oaks, CA: Sage Publications.

Ding, L. (2019). Theoretical perspectives of quantitative physics education research. Physical Review Physics Education Research, 15(2), 020101. doi: 10.1103/PhysRevPhysEducRes.15.020101

Fantini, P. (2014). Verso una teoria locale dell’appropriazione nell’insegnamento/apprendimento della fisica. Doctoral thesis in “Antropologia e Epistemologia della Complessità” Università di Bergamo.

Feuer, M. J., Towne, L., & Shavelson, R. J. (2002). Scientific culture and educational research. Educational Researcher, 31(8), 4–14. http://doi.org/10.3102/0013189x031008004

Fischer, C., Pardos, Z. A., Baker, R. S., ..., & Warschauer, M. (2020). Mining Big Data in Education: Affordances and Challenges. *Review of Research in Education*, 44(1), 130–160.

Glesne, C. (2006). Becoming qualitative researchers: An introduction, Boston, MA: Pearson Education, Inc.

Henderson, C. (2017). Editorial: Call for Papers for Focused Collection of Physical Review Physics Education Research Quantitative Methods in PER: A Critical Examination, *Physical Review Physics Education Research,* 13, 010002. doi: 10.1103/PhysRevPhysEducRes.13.010002.

Henderson, C. (2021). Editorial: Call for Papers for Focused Collection of Physical Review Physics Education Research Qualitative Methods in PER: A Critical Examination, *Physical Review Physics Education Research,* 17, 020001. doi: 10.1103/PhysRevPhysEducRes.17.020001.

Isaac, S., & Michael, W. B. (1995). Handbook in research and evaluation: A collection of principles, methods, and strategies useful in the planning, design, and evaluation of studies in education and the behavioral sciences (3rd ed.). EdITS Publishers.

Jarke, J., & Breiter, A. (2019). Editorial: the datafication of education. *Learning, Media and Technology*, 44:1, 1-6. https://doi.org/10.1080/17439884.2019.1573833

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*, *33*(7), 14-26. https://doi.org/10.3102/0013189X033007014

Klašnja-Milićević, A., Ivanovic, A., & Budimac, M. (2017). Data science in education: big data and learning analytics. *Computer Applications in Engineering Education, 25*, 1068–1078. https://doi.org/10.1002/cae.21844

Krefting, L. (1991). Rigor in Qualitative Research: The Assessment of Trustworthiness. American Journal of Occupational Therapy, 45, 214-222. https://doi.org/10.5014/ajot.45.3.214

Libarkin, J.C., & Kurdziel, J.P., (2002). Research Methodologies in Science Education: The Qualitative-Quantitative Debate, Journal of Geoscience Education. 50 (1), 78-86 https://serc.carleton.edu/files/nagt/jge/columns/ResMeth-v50n1p78.pdf

Mariegaard, S., Seidelin, L. D., & Bruun, J. (2022). Identification of positions in literature using thematic network analysis: the case of early childhood inquiry-based science education. *International Journal of Research & Method in Education*, 45(5), 518-534.

McFarland, D. A., Khanna, S., Domingue, B. W., & Pardos, Z. A. (2021). Education Data Science: Past, Present, Future. *AERA Open* 7. https://doi.org/10.1177/23328584211052055

Odden, T. O. B., Marin, A., & Rudolph, J. L. (2021). How has Science Education changed over the last 100 years? An analysis using natural language processing. *Science Education*, 105, 653– 680. doi: 10.1002/sce.21623

Onwuegbuzie, Anthony & Johnson, R. (2021). The Routledge Reviewer's Guide to Mixed Methods Analysis. doi: 10.4324/9780203729434

Pietsch, W. (2015). Aspects of theory-ladenness in data-intensive science. *Philosophy of Science* 82.5, 905-916. https://doi.org/10.1086/683328

Romero, C., & Ventura, S. (2013). Data Mining in Education. WIREs Data Mining and Knowledge Discovery, 3: 12–27 doi: 10.1002/widm.1075

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, 10 (3). doi: 10.1002/widm.1355

Rossman, G. B., & Rallis S. F. (1998). Learning in the field: An introduction to qualitative research. Thousand Oaks, CA: Sage Publications.

Schutz, A. (1967). Collected Papers I: The Problem of Social Reality (Kluwer Boston, Hingham, MA).

Siemens, G. & Baker, R. S. (2012). Learning analytics and educational data mining: Towards communication and collaboration

Stanley, S. D., & Robertson, W. B. (2024). Qualitative research in science education: A literature review of current publications. European Journal of Science and Mathematics Education, 12(2), 175-199. https://doi.org/10.30935/scimath/14293

Szyjka, S. (2012). Understanding research paradigms: Trends in science education research. Problems of Education in the 21st Century, 43 (1), 110–118. doi: 10.33225/pec/12.43.110.

Taber, K. (2014). Methodological Issues in Science Education Research: A Perspective from the Philosophy of Science. 10.1007/978-94-007-7654-8_57.

Treagust, D.F. & Won, M. (2023). Paradigms in Science Education Research in Lederman, N.G., Zeidler, D.L., & Lederman, J.S. (Eds.). (2023). *Handbook of Research on Science Educatio*n: Volume III (1st ed.). Routledge. https://doi.org/10.4324/9780367855758

Wang, K. D., Cock, J. M., Käser, T., & Bumbacher, E. (2023). A systematic review of empirical studies using log data from open-ended learning environments to measure science and engineering practices. British Journal of Educational Technology, 54, 192–221. https://doi.org/10.1111/bjet.13289

Wise, A. F. & Shaffer, D. W. (2015). Why theory matters more than ever in the age of big data. Journal of Learning Analytics, 2(2), 5–13. http://dx.doi.org/10.18608/jla.2015.22.2

Wulff P., Westphal, A., Mientus, L., Nowak, A., & Borowski, A. (2023). Enhancing writing analytics in science education research with machine learning and natural language processing—Formative assessment of science and non-science preservice teachers' written reflections. *Frontiers in Education*, 7. https://doi.org/10.3389/feduc.2022.1061461

Zhang, J., Wang, W., Xia, F., Lin, y., and Tong, H. (2020). Data-Driven Computational Social Science: A Survey. *Big Data Research,* 21. https://doi.org/10.1016/j.bdr.2020.100145.

Zorić, A. B. (2020). Benefits of Educational Data Mining. *Journal of International Business Research and Marketing, Inovatus Services Ltd.,* vol. 6(1), pages 12-16, November. DOI: 10.18775/jibrm.1849-8558.2015.61.3002

Chapter 2 – Computational AI methods for text analysis

Theoretical overview and case studies

2.1 Introduction

As discussed in Chapter 1, this thesis explores novel approaches to text data analysis composed of two main components: qualitative methods and computational, AI-based methods. Both deserve ongoing reflection to ensure their continued relevance in evolving research contexts. However, qualitative approaches are widely adopted and have already received sustained attention in the science education literature, especially regarding their paradigmatic foundations and methodological implications. In contrast, quantitative approaches emerging from data science and AI are evolving at a remarkable pace, extending well beyond traditional statistical techniques.

Therefore, this chapter explores computational approaches and serves both the reader, and above all the author as an opportunity to immerse in their methodological landscape. First a theoretical overview of these techniques is provided (Section 2.3.), and then two practical case studies are presented.

The first one involves a large-scale corpus from the field of physics education,where I have applied an unsupervised machine learning technique known as Latent Dirichlet Allocation (LDA) to perform a thematic literature review (Section 2.4). LDA belongs to the broader family of natural language processing (NLP) methods developed for textual data analysis. Alongside the specific results of this study, this case study offers a comprehensive overview of the theoretical, practical, and methodological considerations embedded in the analysis process.

The second case study presents a more dialogical example, highlighting the methodological uncertainty researchers often face when choosing between qualitative and quantitative approaches in educational research (Section 2.5). This case originates from a project on future-oriented science education and reflects broader challenges common to studies that analyze student-generated texts to understand perspectives on socio-scientific issues. In this instance, the same corpus was analyzed using both a thematic qualitative method and a computational data-mining technique. This parallel application led to important methodological and epistemological reflections, based on the comparison of the two analytical processes.

Building on this dual engagement with theoretical exploration and practical application, a subsequent reflection distilled a set of core pillars characterising computational AI-based methods in Science Education Research (Section 2.6).

2.2 Aims and research questions

The overall aim of this chapter is to explore computational, data-intensive methods, particularly those that can be employed in SER for textual data analysis. The focus lies in mapping the general methodological features that characterise these approaches. For example, we aim at highlighting what these methods can do and what happens during the analytical process, for example, how text data is treated, the role of researchers, and the kind of methodological choices.

This exploration has been both theoretical and experiential. For a qualitative researcher like myself seeking to expand my methodological repertoire with new tools (before engaging in deeper meta-reflection), both theoretical understanding and practical experience are essential. It would be impossible to grasp why certain approaches work, what epistemic assumptions they embed, or what kinds of questions they are suited to address, without first clarifying the conditions under which these methods operate. Yet, without direct practical engagement, one cannot appreciate their affordances, limitations, or the forms of understanding they enable in real research contexts. Both dimensions are therefore crucial for situating these methods within their broader theoretical landscape and for addressing questions such as: "What are their basic characteristics?", "what new types of knowledge do they allow us to generate?", "what are the main strengths?", and "what boundaries define their effective use?".

Accordingly, this chapter accompanies the reader through my personal encounter with computational and AI-based methods for text analysis in SER. Through this experience, I aim to transform what began as an individual learning process into a shared reflection for the research community, highlighting insights, difficulties, and considerations that may resonate with other scholars navigating similar methodological transitions.

We begin with a theoretical exploration, positioning these methods conceptually before turning to their practical implementation. This mirrors the process I followed as a researcher: before using such methods, I first built awareness of their functioning and of the theoretical principles on which they rely.

Specifically, this chapter is guided by the following research questions:

1. What are the theoretical foundations of computational AI-based methods, and what are the pillars on which they rest?
2. What practical challenges emerge when applying them to textual data in SER and what benefits or new opportunities do these methods offer for SER?

2.3 Framing the landscape of computational methods

2.3.1 The hybrid birth of Artificial Intelligence

As Jean-Gabriel Ganascia argues in *Epistemology of AI Revisited in the Light of the Philosophy of Information* (2010), AI was born as a profoundly hybrid epistemic project. From its origins at the Dartmouth Conference (1956), AI integrated two seemingly divergent traditions: on the one hand, the symbolic and rationalist tradition, rooted in logic and mathematics, which sought to represent cognition through formal rules and symbolic manipulation; on the other, the biological and connectionist tradition, inspired by cybernetics and neural systems, which aimed to reproduce adaptive mechanisms of perception and learning.

This dual origin gave AI a distinctive position between the sciences of nature and the sciences of culture. The first tendency approached intelligence as a natural phenomenon describable through general laws and inductive logic; the second viewed intelligence as a cultural, goal-oriented activity comprehensible only by analyzing intentions, contexts, and representations. In Ganascia's terms, AI belongs to an intermediary domain, simultaneously theoretical, empirical, and interpretive: it constructs formal models of reasoning (as the natural sciences do) while engaging with meaning and context (as the humanities do).

This hybridity has deep methodological implications. It shows that the opposition between qualitative and quantitative approaches is historically misplaced: AI has always operated at their intersection. Symbolic reasoning and statistical learning, far from being incompatible, represent the two poles of a single epistemic system, one that aspires both to formalization and to interpretation. Describing contemporary computational methods as mixed is therefore not a novelty, but a return to AI's original condition: a science of modelling that simultaneously analyzes structure and meaning.

Seen in this light, the persistent ambiguity surrounding computational analyses in SER, *"are they qualitative, quantitative, or both?"*, echoes the duality at the core of AI itself. The field was conceived as an experiment in bridging logics: between deduction and induction, between the artificial and the cultural, between the symbolic and the embodied. Recognizing this lineage helps us frame present-day computational methods not as epistemic anomalies but as heirs to a long tradition of hybrid reasoning.

Ganascia's interpretation of AI as an "intermediary science" offers a conceptual bridge to the methods examined in this chapter. If AI constitutes a science of the artificial, a domain where knowledge is produced through artefacts that mediate between the natural and the cultural, then each contemporary technique can be understood as a model of knowledge. Methods such as topic modeling, natural language processing (NLP), and large language models (LLMs) are not neutral tools for text analysis: they are epistemic artefacts that instantiate particular ways of representing and manipulating knowledge. Topic modeling, for instance, represents texts as probabilistic mixtures of latent themes; NLP techniques translate words into numerical vectors that encode semantic proximity; LLMs generate language by predicting probable continuations within learned distributions. Each of these operations embodies a theory of meaning, a computational metaphor of how knowledge can be structured, related, and reproduced.

In SER, these methods function not merely as instruments for analyzing student texts, but as mediators between researchers and knowledge phenomena. They offer representations of reasoning and meaning that are different from, but increasingly intertwined with, traditional qualitative interpretation. In this sense, the researcher no longer confronts data directly but engages in a dialogue with the model, interpreting its outputs as one interprets the readings of an experimental apparatus. As in the natural sciences, where scientists must trust instruments that extend their perceptual reach while remaining aware of the models that underpin them, researchers employing AI-based methods must develop an awareness of the assumptions, transformations, and boundaries that define what these models can, and cannot, reveal about human meaning-making.

2.3.2 Computational models for text analysis: theoretical overview

The following overview does not aim for technical completeness but rather for *conceptual orientation*. The purpose here is to sketch the main families of computational techniques that are central to textual data analysis, both in general and, by extension, in SER. These short descriptions are intended to convey the logic of each model, rather than to provide detailed definitions or exhaustive accounts of their mechanisms, which would exceed the scope of this chapter. What matters for the present reflection is that each computational method embodies a distinct way of transforming language into data, and of representing meaning in a form that can be analysed, compared, or modelled.

Artificial Intelligence has been defined in multiple ways, reflecting its inherently plural and evolving nature. Russell and Norvig (1995, p. 5) summarising eight different definitions, refer to AI as "systems that think like humans, systems that think rationally, systems that act like humans, and systems that act rationally". Each of these perspectives implies a different understanding of intelligence and of the relationship between cognition, representation, and action. In this chapter, however, AI is considered broadly as the set of computational systems or *models* that think and act

rationally, that is, systems designed to reason from data and to act upon such reasoning through formal operations. When applied to text, these models instantiate specific ways of representing linguistic meaning in computational form, making language analysable through patterns, probabilities, and relations.

This view aligns with what Paul Humphreys (2004) called the *computational extension of ourselves*: the idea that models are epistemic artefacts that expand human capacities for perception and reasoning. In this sense, adopting AI-based methods in educational research entails working within a form of *mediated empiricism*, where observation and interpretation are channelled through artefacts whose assumptions and abstractions shape what can be known. Recognising this model-based nature of AI is essential, because it positions computational methods not merely as technical instruments, but as epistemic agents that participate in the construction of knowledge. Similarly, Krist and colleagues (2025) discuss the concept of *AI as IA*, that means Intelligence Augmentation, the idea that computational systems should be understood as tools that extend and support human reasoning rather than replace it. From this perspective, machine learning and related methods are most valuable when they augment interpretive and analytic capacities, helping researchers detect patterns or relations that might otherwise remain unnoticed, while leaving the work of inference, meaning-making, and theoretical interpretation firmly in human hands. This reframing emphasizes the importance of using AI critically and contextually, as part of a human-machine partnership aimed at enhancing, not automating, knowledge production.

The following list briefly outlines the main families of computational models that have become central to text analysis. Each is presented not for its technical detail, but for the specific way it translates linguistic meaning into computable form and thereby shapes what kind of knowledge can be produced.

- **Machine Learning:** Machine Learning (ML) learns from examples to generalize to new cases. In ML, knowledge is operationalized as patterns of correlation discovered across datasets. The epistemic logic is primarily inductive: models infer associations that support prediction rather than uncovering causal mechanisms. This brings immense scalability and sensitivity to subtle regularities, but it also narrows the observable world to what is frequent, measurable, and computationally tractable. Rare, ambiguous, or context-dependent phenomena, central in educational discourse, are at risk of being filtered out unless deliberately modeled.
- **Natural Language Processing (NLP):** NLP provides the toolkit for transforming natural language into analyzable numerical data. Tokens (words, subwords) become features; documents become vectors or matrices. Classic methods include bag-of-words and TF-IDF representations, n-grams, part-of-speech tagging, parsing, and basic semantic features. The transformational principle is key: to make text computable, models deliberately reduce complexity, preserving some aspects of meaning while discarding others (e.g., bag-of-words ignores word order). This is not a flaw but a modelling choice with epistemic consequences.:
- **Word embeddings and vector models of meaning (e.g., Word2Vec):** Vector models map words (or larger units) into continuous spaces where geometric proximity encodes contextual similarity. They excel at capturing relational semantics (“king − man + woman ≈ queen”) but do not by themselves encode conceptual, pragmatic, or situated meaning (Mikolov, et al., 2013). A tacit assumption often at play is a variant of “meaning is use”: distributional regularities in language serve as proxies for cognitive or conceptual relations. This is powerful for large-corpus regularities and weaker for unique, context-bound interpretations.

- **Topic Modeling as a probabilistic representation of themes:** Topic models such as LDA treat each document as a mixture of themes, each theme as a distribution over words. Knowledge is represented as distributions rather than certainties, making topic modeling well-suited to mapping large corpora and identifying emergent structures. The trade-off is a filtering effect: what is infrequent or weakly coherent may be suppressed by the model's assumptions, which privilege recurring patterns over singular, context-specific nuances.
- **Network Analysis and relational models**: Network approaches conceptualize texts (or concepts, authors, terms) as nodes connected by relations (co-occurrence, citation, semantic links). Meaning is treated as an emergent property of structure, centrality, communities, paths, rather than as an attribute of isolated units. This relational ontology aligns with complex systems thinking and is particularly illuminating when the goal is to understand how ideas cluster, diffuse, or stabilize across communities and time. Networks can be informative even with moderate data sizes, though larger graphs reveal more robust mesoscale patterns.
- **Large Language Models (LLMs):** LLMs are deep, generative architectures trained on massive corpora to predict the next token. They model probabilistic co-occurrence in language and can simulate discourse, summarize texts, or generate analyses that appear strikingly fluent. Their epistemic object, however, is linguistic plausibility, not truth: they optimize coherence within learned distributions. This makes them powerful assistants for pattern-oriented tasks and risky sources for unique facts or context-heavy interpretation unless used with explicit safeguards and human oversight.

All these models and concepts will reappear, in different forms, throughout this thesis. They constitute the conceptual and methodological repertoire through which the subsequent studies engage with text as a site of knowledge construction in Science Education Research. By encountering these models in practice, we will be able to examine more concretely how they operate, what kinds of insights they make possible, and how they reshape the researcher's relation to data and meaning.

2.3.3 The lesson of hallucinations: why understanding models matters

In this final section of the theoretical exploration, we bring LLMs into view to illustrate why understanding foundational assumptions is essential for understanding what the model can do in an excellent way, but also where it is prone to fail. This is fundamental to be aware of how humans and machines can work together and also play a virtuous role in scientific knowledge production, maintaining scientific trustworthiness.

The recent debate on hallucinations in large language models offers a paradigmatic example. Far from being random errors or mere artefacts of technical limitations, hallucinations reveal the inner logic of these systems. As demonstrated in the OpenAI study *Why Language Models Hallucinate* (Kalai, Nachum, Vempala, & Zhang, 2025), they emerge because the objectives used to train and reward these models structurally favour confident guessing over epistemic humility. In common training pipelines, models are rewarded for producing specific answers rather than for admitting uncertainty. When evaluation is effectively binary, "correct" earns credit, "incorrect" or "I don't know" does not, the rational strategy for the model is to guess with confidence, even at the cost of truth.

When reward structures are modified to grant partial credit for appropriate uncertainty and to penalize overconfident errors more than abstentions, models respond with far more "I don't know" answers and substantially fewer hallucinations. The key insight is that apparent overconfidence is not a psychological flaw but an emergent property of the model's epistemic environment, the training and evaluation rules that define success.

Crucially, hallucinations concentrate on tasks that lack stable patterns, such as remembering dates of birth, specific historical dates, or other isolated facts. These are "single-instance" data points; they do not form statistical regularities from which the model can generalize. LLMs are designed to learn patterns, repeated co-occurrences and probabilistic associations in language. They thrive on regularities, analogies, and distributions, not on unique or context-specific facts. Confronted with objects that have no learnable structure, like a date that appears once in the training corpus, the model extrapolates from linguistic proximity, effectively inventing coherence where none exists.

This limitation is epistemological. LLMs do not know in the human sense; they model the likelihood that one sequence of words follows another. Their "knowledge" is statistical rather than semantic. They are optimized for plausibility, not truth. Thus, hallucinations are not malfunctions but faithful expressions of the epistemic commitments built into their architectures and training regimes.

For researchers in Science Education, the lesson is broader. The "black box" of AI is not merely a technical opacity but a space of epistemic negotiation. It materializes decisions about what counts as evidence, coherence, and uncertainty. Understanding these assumptions becomes crucial when interpreting AI-generated analyses of human reasoning or discourse. As experimental physicists must understand how instruments transform phenomena into observable signals, so educational researchers must read algorithmic outputs as *model-dependent projections*. In other words, AI outputs are not direct reflections of reality, but structured and manipulated projections of it.

2.4 First case study: LDA and literature review in physics education

2.4.1 Rationale and context

This first case study shows the application of computational methods to the analysis of textual data within the field of Physics Education Research (PER). It represents a concrete example of how computational, AI-based methods can be used for textual data analysis in our field. Specifically, it focuses on the use of *Latent Dirichlet Allocation* (LDA), an unsupervised topic-modelling technique, within the broader family of NLP methods, to perform a large-scale thematic analysis of the physics education literature.

The study began during my PhD research stay at the University of Oslo's Center for Computing in Science Education (CCSE), where I worked under the supervision of Tor Ole B. Odden. This setting offered an ideal context in which to explore the methodological potential of data-intensive approaches to science education. The CCSE is a leading institution in developing computational and data-driven methods for understanding science learning, and it was within this vibrant environment that I could directly experience the integration of human and computational reasoning in research practice.

During my stay, I began by exploring a dataset composed of all articles coming from *The Physics Teacher* (TPT) journal from its inception to 2020. This preliminary study allowed me to familiarise myself with the model, understand the nature of the analysis, and obtain initial insights into how topic modelling could be used for literature reviews in the context of physics education. A first version of this work was presented at the 4th World Conference on Physics Education (WCPE 2024) and published as a conference proceeding. Following the review process, our paper "*Reviewing 55 years of Literature on Physics Education using Natural Language Processing"* (Caramaschi, Odden, & Levrini) was selected for inclusion in the Springer Book of Selected Papers, part of the *Challenges in Physics Education* series.

Building on this preliminary exploration, we extended the work beyond a literature review to compare two different systems of scholarly production, the North-American and European traditions in physics education. Thus, we analysed two large-scale corpora from *The Physics Teacher* and *Physics Education* journals. This subsequent stage, presented in this chapter, represents the mature development of that research line and is documented in *Analyzing the History of Physics Education in the USA and Europe through Natural Language Processing* (Caramaschi & Odden, 2025). The analysis spans more than five decades of literature, providing a long-term view of how themes in physics education evolved across both American and European contexts.

From a reflective perspective, the inclusion of this case study serves as an auto-ethnographic account of methodological learning. Through hands-on engagement with computational modelling, I gradually developed an understanding of both the affordances and the boundaries of machine-learning techniques for textual analysis. The process of reproducing, refining, and extending earlier work transformed what initially began as a methodological exercise into a reflective inquiry into what it means to interpret educational knowledge through the lens of computation. The case thus stands not only as empirical evidence but also as a lived example of methodological evolution, a bridge between theoretical awareness and practical expertise in computational approaches to Science Education Research.

2.4.2 Aims and Research question of the study

While previous studies have successfully employed Latent Dirichlet Allocation (LDA) to examine thematic trends in science and physics education research, most have focused on *research-oriented* literature rather than *teaching-focused* journals. Moreover, earlier analyses have typically investigated either a single venue such as with articles from *Science Education* (Odden et al., 2020) or conference proceedings, for instance the *Physics Education Research Conference* series (Odden et al., 2021). As a result, there remains a gap in our understanding of how educational discourse has evolved across journals specifically aimed at physics educators and practitioners.

This study addresses that gap by applying LDA to two long-standing, internationally recognized journals: *The Physics Teacher* (TPT) and *Physics Education* (PE). Both journals share a commitment to advancing physics teaching but differ in audience, geographical orientation, and editorial scope. TPT, published by the American Association of Physics Teachers, reflects the North American context, emphasizing classroom innovation and instructional practice. PE, published by the Institute of Physics, offers a European perspective, often focusing on curriculum development, conceptual understanding, and didactical theory.

By examining the thematic evolution of these two journals over more than five decades, this study aims to reveal how priorities in physics teaching and learning have changed over time and across educational systems. It seeks to connect these developments to broader methodological and epistemological trends, thereby situating the findings within the historical and theoretical reflections introduced in Chapter 2. In this sense, the case study functions not only as an empirical investigation but also as a continuation of the methodological reflection that frames this thesis.

The research questions that guided the literature review were as follows:

1. What topics can be discovered in the corpus of articles composed of all articles published in *The Physics Teacher* and *Physics Education* journal from 1966 through 2019? What

characteristic trends about physics teaching can be recognised, how did they evolve, and what is their prominence in this ensemble of articles?
2. To what extent the discovered topics are covered by the two journals separately? Are there specific trends or interests that differentiate the two journals?
3. Considering the reference literature, what do these topics and their trends tell us about the differences and similarities between the two educational systems (European and North American) in regard to the teaching of physics?

2.4.3 Latent Dirichlet Allocation model

Theoretical overview of LDA

To understand the methodological core of this case study, it is useful to begin from the model itself. This section presents the theoretical principles underlying Latent Dirichlet Allocation (LDA) and explains how its structure and assumptions resonate with the epistemological reflections developed throughout this thesis.

LDA is a probabilistic generative model for *topic modelling*, first introduced by Blei, Ng, and Jordan (2003). Developed within the field of Natural Language Processing (NLP), a branch of artificial intelligence concerned with the computational representation of language as introduced in *Section 2.3,* LDA has since become a foundational tool in data-intensive research across the humanities and social sciences for its ability to identify topics in textual corpus. LDA belongs to the family of unsupervised machine-learning methods, meaning that it uncovers latent structures within data without pre-labelled categories. It does not "know" what the topics are in advance. Instead, it infers topics from statistical regularities in how words co-occur across a corpus. In this sense, it operates as a *data-driven model of discovery*, designed to reveal patterns that emerge from the internal organization of the data rather than external theoretical imposition.

The model assumes that documents are mixtures of topics, and that topics are mixtures of words. Formally, this is represented through two distributions:

- a topic distribution over documents: each document expresses multiple topics, each with a certain probability weight;
- a word distribution over topics: each topic is characterized by a specific set of words that are likely to appear together across documents.

In this way, LDA provides a two-layer generative representation of textual meaning, which can be understood as a simplified *model* of the writing process. When constructing a document, an author typically combines multiple themes or lines of reasoning. LDA formalises this intuition by modelling a text as a combination of several latent topics, and each topic as a probabilistic mixture of words that tend to occur together. In this abstraction, writing is represented as the act of assembling topics in varying proportions, and each topic, in turn, is represented by a characteristic pattern of words that co-occur across the corpus. The model then reverses this hypothetical writing process: given only the words present in the documents, it infers the hidden topic structure that could most plausibly have generated them. LDA does not reproduce how humans actually write; rather, it offers a computational analogy for how meaning can be represented as layered combinations of linguistic regularities.

This dual distributional structure expresses an epistemic idea deeply aligned with the data-intensive paradigm: that meaning can be represented as a set of probabilistic relations among linguistic

elements. Rather than treating words as fixed semantic entities, LDA treats them as *signals of association*. Co-occurrence becomes the proxy for conceptual relation, and meaning emerges statistically from repetition and proximity.

Technically, LDA operates on a bag-of-words (BoW) representation of text, where the order of words is disregarded and only their frequencies are retained. This simplification allows the model to handle large corpora but also entails an epistemic transformation: language is reduced to a distributional structure that preserves *how often* words appear together, but not *how* they are connected syntactically or semantically. Through iterative inference (usually via Gibbs sampling or variational Bayes), the algorithm identifies groups of words that co-occur across many documents more frequently than would be expected by chance. Each group becomes a *topic*, represented by a ranked list of words with associated probabilities. The researcher then interprets these lists to assign thematic labels and to trace topic prevalence across time or across different sources.

At a conceptual level, the model can be envisioned as a network of relations linking three layers: *words*, *topics*, and *documents*. Every document is described as a blend of topics, and each topic is in turn defined as a weighted combination of all the words that appear in the corpus. The entire collection of documents determines the dictionary, which is the complete set of unique words that the model recognises and uses to build its internal structure. Importantly, every topic contains all the words of this dictionary, but each word has a different *probability weight*. Words that frequently co-occur in similar contexts have higher weights within the same topic, while others remain present with minimal values. The distinctiveness of a topic, therefore, does not depend on which words it includes, but on the *pattern of probabilities* it assigns to them. Two topics may share many words, yet differ profoundly in how those words are emphasised.

Through this architecture, LDA establishes a bidirectional link between the micro level of language and the macro level of documents: topics connect words that tend to appear together, and documents connect topics that tend to co-occur in similar texts. This allows researchers to move between levels, observing how specific terms shape broader themes, and how these themes, in turn, define the intellectual structure of a corpus.

From an epistemological standpoint, this structure implies a specific *model of knowledge*, that is relational, probabilistic, and distributed. Knowledge is not located in any single document, but in the patterns that connect documents through shared linguistic features. Similarly, a concept or "topic" is not defined by a single term but by a constellation of co-occurring expressions whose statistical coherence hints at an underlying theme.

Interpretation and researcher's agency

Despite its mathematical precision, LDA does not produce "topics" in the semantic sense. It generates word distributions whose interpretation depends on the researcher's domain knowledge and theoretical sensitivity. This interpretive act is essential: researchers must decide how many topics to extract, how to preprocess the corpus, and how to interpret each cluster of words in context.

Such decisions instantiate what was previously defined as model-dependence, the idea that every computational representation of knowledge embodies specific assumptions and boundaries of validity. Choices like the number of topics, the strength of topic overlap (controlled by Dirichlet priors α and η), or the inclusion of bigrams versus single tokens all influence the structure of the results. Moreover,

because LDA is probabilistic, repeated runs may yield slightly different configurations; ensuring stability thus requires both computational and interpretive validation.

For this reason, the meaning of an LDA model cannot be separated from the researcher's modeling decisions and reflective engagement. The machine identifies possible regularities within the conditions set by those decisions, while the human interprets them as epistemically meaningful patterns. The dialogue between the two constitutes the essence of human–machine complementarity in data-intensive educational research.

Why LDA suits this thesis

In this thesis, LDA is not conceived as a neutral tool for classification but as a computational model of knowledge, a method that embodies particular epistemic commitments. It operationalises transformational reduction, converting rich textual meaning into probabilistic form; it captures relational structures rather than causal chains; and it foregrounds model-dependence, reminding us that computational insight always emerges from deliberate design.

LDA was therefore chosen for both practical and conceptual reasons. Practically, it allows the exploration of large, multi-decade corpora that are beyond the reach of traditional qualitative methods. Conceptually, it provides a model through which to reflect on how knowledge is represented, abstracted, and reinterpreted by computational systems. The interpretive work that follows, then, does not simply apply LDA, but *dialogues with it*, treating the model as a partner in inquiry rather than a mechanical instrument.

2.4.4 Overview of the analytical process

In this case study, two analyses were conducted to address the research questions. First, an inductive topic analysis using LDA on the combined corpus of *The Physics Teacher* (TPT) and *Physics Education* (PE), in order to identify the latent themes that characterise the literature across both journals. Second, a comparative quantitative analysis of these results, distinguishing TPT and PE articles to examine potential differences between the two journals and the communities they represent.

Even though the two analyses are presented here as logically sequential, the research process itself unfolded in a more complex and exploratory manner. Before performing the inductive analysis on the complete dataset (and before even conceiving the study as a comparative one) we began with a more open and experimental attitude. Although LDA had already proven effective in previous studies, it was not immediately clear that it would yield meaningful or original results on this type of textual data, given its diversity of genres and publication contexts.

A key step in this early phase involved exploratory LDA analyses run separately on each journal. These preliminary models had a heuristic purpose: they helped familiarise us with the data, informed technical choices (such as the range of topic numbers), and guided calibration of preprocessing steps. Both single-journal analyses suggested that around fifteen topics meaningfully described each corpus, which in turn guided the minimum topic number explored in the combined dataset. The analysis on TPT corpus was developed as part of my preliminary work (Caramaschi, Odden, & Levrini, *in press*), which laid the foundation for the broader comparative study presented in this chapter.

In the following sections, this process is described in detail, from data selection and data collection and preparation to model training, interpretation, and validation. Methodological reflections

accompany each phase, highlighting how practical and epistemic considerations intertwined throughout the research process.

2.4.5 Detailed methodology

Each methodological step presented below is described in full detail in the aforementioned research article *"Analyzing the History of Physics Education in the USA and Europe through Natural Language Processing"* (Caramaschi & Odden, 2025). The following summary provides an overview of the key procedures and methodological reasoning most relevant for the purposes of this thesis.

Journal selection rationale

The corpus analysed in this study consists of all published articles from two long-standing journals: *The Physics Teacher*[1] (AAPT/AIP, USA) and *Physics Education*[2] (IOP, UK). These journals have been published since 1963 and 1966 respectively, providing a unique window into the evolution of physics teaching discourse across North America and Europe. For the purposes of this study, the analysis focuses on the overlapping timeframe between 1966 and 2019, which ensures full comparability between the two sources.

From a methodological standpoint, selecting these two journals was not only a pragmatic decision but also an epistemic one. It reflected an effort to build a corpus capable of representing the pedagogical discourse of physics teaching in two major cultural and institutional systems. The focus on English-language publications ensured accessibility and comparability across contexts, but also introduced the need for critical reflection on representation and inclusivity.

Although other national journals on physics teaching exist (e.g., *Il Giornale di Fisica*, *Fra Fysikkens Verden*, *Phydid A/B*), these are typically non-English and locally focused. The choice of *TPT* and *PE* was therefore guided by their international reach, historical continuity, and thematic comparability, features that made them suitable for a cross-regional analysis of educational discourse.

For the coverage and nature of the journal, we were confident that *TPT* could serve as a reliable representation of the North American context, but the European situation required further examination. To validate this assumption, we evaluated the geographical representativeness of *PE* by analysing author affiliations for selected years (1966, 1980, 2000, 2010, 2019) using the OpenAlex dataset. Results indicated that early publications were predominantly European (mostly UK-based), while more recent decades show increasing global participation (≈ 53% Europe, 20% North America, 16% Asia, 8% Oceania).

This validation step exemplifies the reflexive attitude that guided the entire methodological process: representativeness was not taken for granted but empirically examined. The resulting insight, that *PE* evolved from a predominantly European to an international outlet, was incorporated into the interpretive framework, allowing later results to be contextualised in relation to the changing identity of the journal itself.

[1] https://pubs.aip.org/aapt/pte
[2] https://iopscience.iop.org/journal/0031-9120

Dataset creation and text preparation

The creation of a coherent and analyzable corpus was the first crucial step in the analysis process. This phase involved not only the mechanical conversion of documents into machine-readable form but also a series of interpretive choices that shaped the analytical scope of the study. All available articles were downloaded in PDF format and converted into text using Adobe Acrobat, with metadata such as title, year, authors, and DOI stored in structured tables. Because the *TPT* and *PE* datasets were retrieved four years apart (2020 and 2024), small procedural differences arose due to advances in text-extraction tools, as detailed below.

The initial curation phase aimed to ensure that only content relevant to the study's objectives was retained. Editorials, announcements, advertisements, book reviews, and other non-substantive materials were excluded, along with very short documents (< 400 characters) and those lacking author information. A qualitative inspection of both datasets revealed recurring series (e.g., *Apparatus for Teaching Physics*, *Notes*, *Research Frontiers* in TPT; *Letters*, *People*, *Reviews* in PE). Decisions on inclusion or exclusion were made by considering their relevance to the research questions, ensuring that the corpus captured disciplinary discourse rather than editorial or institutional communication.

A second round of cleaning ensured consistency between journals and removed redundant or noisy data (e.g., *Corrections*, *Errata*, *Book Reviews*, *News*). Short texts (< 500 words) were excluded, on the assumption that longer articles offer richer thematic material. The final corpus comprised 13,648 articles (7,203 from TPT and 6,445 from PE), providing a robust basis for comparative analysis.

The dataset presented some limitations, primarily related to *TPT*'s older articles, that often contained OCR errors due to scanned sources, such as incorrect merging of sections or misplaced words. An inspection of a random sample of articles revealed that the issues were most pronounced in those published before 1978. However, in a bag-of-words model such as LDA, the order of words in a text is not a crucial factor. Consequently, we were confident that occasional merging errors do not significantly affect model validity, a greater reason in our case, in which short articles were excluded from the dataset, leaving only longer documents where small textual anomalies are unlikely to distort overall word-usage patterns.

At this stage, consistency across documents was ensured and the dimensionality of the data was reduced. For example, unique identifiers and common section headings (e.g., *Introduction*, *Methodology*, *Results*) were removed, along with newline characters, tabs, digits, and bullet points. Non-alphanumeric symbols were replaced with spaces, hyphenated fragments at line breaks were merged (e.g., "pre-\nprocessing" becomes "preprocessing"), and all text was converted to lowercase. Minor lexical corrections were also applied, such as fixing frequent OCR typos ("tlie*"* becomes "the", "per cent" becomes "percent"). These steps yielded a clean and standardised text suitable for computational analysis.

Finally the text was prepared for modeling: sentences were tokenised into lowercase word sequences while removing accent marks and punctuation. Stopwords (e.g., "if", "and", "but") were removed to retain only semantically relevant terms. Lemmatization was performed using WordNet, assigning each token its most likely part-of-speech tag (e.g., "matches" becomes "match") before reducing it to its base form. Finally, bigrams, which are frequent co-occurring word pairs, were generated to preserve multi-word expressions (e.g., "high school" becomes *high_school*), enhancing contextual understanding in the subsequent analysis.

This is an overview of how the preparation of the dataset happened, combining considerations that took into account contextual aspects such as which words to include or not, but also aspects of the model, such as the impact or advantages brought by a BoW representation and how the model would extract the topics that sometimes led us to discard texts that were too short, always guided by the research questions.

LDA model development and testing

After preparing the textual datasets for *TPT* and *PE*, the combined corpus was used to develop the LDA model. Several parameter-tuning steps ensured robust and meaningful topic extraction.

- **Word frequency filtering:** Before applying LDA, words were filtered based on their frequency. Terms appearing fewer than 15 times were removed to eliminate noise, while overly common terms (occurring in 45-99% of documents) were excluded to avoid trivial patterns. Thresholds were iteratively tuned to preserve important disciplinary terms such as "data", "force", "model", "laboratory", "measurement", and "motion".
- **Model configuration and validation.** The model was implemented using the open-source *Gensim* library (Blei et al., 2003; Řehůřek & Sojka, 2010) in combination with *NLTK* for linguistic processing. The number of topics was tested within the range of 15-22, guided by previous single-journal analyses. Topic coherence was computed to assess interpretability, with typical values for academic corpora between 0.4 and 0.6. The final model, comprising 20 topics, achieved a coherence score of 0.577, slightly above the mean of tested configurations.
- **Stability and reliability:** Because LDA is probabilistic, different random initialisations can yield slightly varied topic structures. Following Odden, Marin, and Caballero (2020), fifteen models were trained and clustered to identify the configuration closest to the mean solution. This model showed high stability and above-average coherence, confirming the reliability of the final topic distribution.

Topic interpretation and trend visualisation

Once the final LDA model was selected, we conducted a face validity check by manually inspecting the most representative words for each topic to ensure coherence and interpretability. The trained model was then applied to the entire set of articles to classify each one according to its distribution across all topics. In practice, each article was assigned a probability value for every topic, reflecting the proportion of its content associated with that theme.

Based on these distributions, we identified the 15 documents most strongly associated with each topic (typically ranging between 60% and 98% topic weight). These highly representative articles, together with the top-weighted words, were used to assign meaningful and contextually grounded labels to each topic. For example, if an article is 89% representative of Topic 1, it means that 89% of its content is classified as belonging to that topic, while the remaining 11% is distributed across other topics. These exemplar articles played a central role in interpretation, providing concrete textual evidence that contextualised each theme and clarified the type of contributions it encompassed.

Topic labels were thus assigned by combining the most representative words with the content of these exemplar articles. Labels were chosen to be concise and meaningful within the context of physics and physics education: general labels were used for broad themes (e.g., *Newtonian Mechanics*), whereas more specific ones were assigned when a topic focused on a particular aspect (e.g., *Trends in Physics Education: Course Structures and Career Pathways*).

Finally, we explored temporal trends to examine how the relative prominence of topics evolved over time. For this analysis, annual topic prevalence was calculated by averaging topic contributions across all documents published each year. By "prevalence," we refer to the proportion of the literature represented by a given topic within a specific year. To smooth fluctuations, we applied a three-year rolling average, preserving macro-trends while reducing statistical noise.

The analysis covered the interval 1966-2019, corresponding to the overlapping publication timeframe for both journals. The endpoint, 2019, was selected because the *TPT* dataset was collected in early 2020, and only fully completed years were included. This timeframe allows the observation of thematic developments from the late 1960s up to the pre-pandemic period, just before the onset of significant transformations such as the rise of large-scale e-learning initiatives and the introduction of AI-based tools in educational practice.

These methodological choices ensured that the LDA model produced interpretable, reliable, and epistemically meaningful topic distributions. The resulting outputs were subsequently analysed to construct narratives about long-term trends in the physics education literature, linking computational findings with interpretive understanding. To ensure transparency and reproducibility, all materials that support the findings of this study are openly available. Specifically, the article metadata, topic weights, and bag-of-words text representations (including bigrams) are publicly accessible via the Zenodo dataset[3] and the GitHub repository[4]. The original article texts themselves cannot be made publicly available due to copyright restrictions.

2.4.6 Literature reconstruction: history of physics teaching in EU and US

This section presents an interpretive reconstruction of the history of physics education in the United States and Europe. The reconstruction builds on the one in the published article, where the analysis aimed to thematically investigate literature on physics education from the 1960s to the present. To contextualize the themes that emerged from the use of LDA within a qualitative study, it was both useful and necessary to examine the significant events and shifts that shaped the evolution of physics education during this period.

The historical overview was developed primarily on the basis of two key sources: Meltzer and Otero (2016), which offers a detailed account of developments in U.S. physics education since the nineteenth century, and Lewis (1999), the chapter from *125 Years: The Physical Society and The Institute of Physics*, which traces the evolution of physics education in Europe. Together, these sources provide complementary perspectives on how political, institutional, and cultural factors shaped the field in different contexts.

From a methodological perspective, this reconstruction functions as a contextual framework for interpreting the thematic patterns identified by the LDA model. Understanding the historical trajectory of physics education was crucial both for naming the topics and for interpreting how clusters in the model correspond to broader transformations in teaching and learning practices. Historical awareness thus becomes part of the analytical process itself, embodying the complementarity between human interpretation and computational modelling that underpins the methodological perspective of this thesis. In this section we decided to retrace the history of physics education by dividing it into temporal periods that I identified as significant, of which it is also possible to see a summary table in

[3] https://zenodo.org/records/17190057
[4] https://github.com/martinacaramaschi/TPT-PE-thematic-analysis

Table 2.1. Also, for each period, a column shows interpretative insights, to make the reader understand how this "external knowledge" was injected into the process as interpretative knowledge.

Table 2.1 - Chronological synthesis of key developments in physics education across the United States and Europe. The table combines historical events with interpretive insights used to contextualize and validate the LDA results, illustrating how transformations in educational practice and discourse resonate with thematic structures identified computationally.

Period	Key events	Interpretative insights
1950s	● Post WWII ● Cold War ● Sputnik launch ● CERN foundation (1954) ● National Science Foundation ● PSSC ● Development of high school and university physics curricula ● General stagnation in European physics education	These developments correspond to the early *curricular reform* themes captured by LDA, showing the dominance of vocabulary linked to curriculum, experimentation, and teaching reform in the early decades of *The Physics Teacher*.
1960s - 1970s	● Project Physics (1962) ● Conceptual and inquiry-based approach to learning. ● Focus on teacher training ● Development of teaching material ● Science for All at CERN ● Diffusion of Nuffield Science (and Physics) Project (1957)	The *teacher training* and *inquiry* topics found in early issues of *Physics Education* reflect these transatlantic currents of reform, highlighting how global the discourse on modernization had become.
1980s	● Establishment of Physics Education Research as academic discipline ● Studies on students' learning difficulties ● Force Concept Inventory (1985) ● Shortage of Physics teachers in Europe ● 4-years degrees in Physics	In the model's temporal dynamics, this phase corresponds to the emergence of topics related to *conceptual change*, *students' reasoning*, and *diagnostics*, evidencing how methodological innovation in PER became a defining axis of the field.
1990s	● Conceptual physics and qualitative understanding over formalism ● Rise of Computational physics ● Simulators included into physics education ● Integration of PER findings into classroom practice	The prevalence of terms in the model related to *technology*, *simulation*, and *learning environments* aligns closely with this era's pedagogical modernization and the expansion of computational literacy in physics education.

2000s - 2010s	• Student-driven exploration • Innovation in lab design: computational tools and interactive, technology-integrated physics labs • Focus on students skills (e.g., troubleshooting skills) • Results from research as a guide for improving instruction	In the LDA model, topics connecting *representation*, *skills*, and *technology* rise steadily after 2000, mapping onto this global wave of laboratory innovation and student-centered reform. Also, this expansion toward interdisciplinarity and reflection is mirrored in our model's appearance of hybrid clusters linking *argumentation*, *students' reasoning*, and *representation*, signalling a mature and self-reflective phase of the discipline.

The reconstruction starts from the 1950s, a period that could be resumed as "postwar expansion and curriculum reform". In fact, the postwar decades marked a period of unparalleled political, scientific, and educational change. In the United States, the Cold War, the establishment of the National Science Foundation (1950), and the launch of *Sputnik* (1957) created an environment where science education was framed as a matter of national security and cultural prestige. Physics, in particular, became a strategic field for demonstrating scientific leadership. Programs such as the Physical Science Study Committee (PSSC, 1956) and Project Physics (1962) exemplified the new orientation toward conceptual, inquiry-based learning, emphasizing experimentation and problem-solving rather than rote memorization. These reforms also represented an early instance of large-scale collaboration between physicists, educators, and psychologists.

In Europe, the 1950s were characterized by relative stagnation in physics education, with teaching often centered on formalism and memorization. Nonetheless, important seeds of change were being planted. The founding of CERN in 1954 symbolized Europe's renewed commitment to scientific collaboration and technological innovation, indirectly inspiring educational modernization. Initiatives like the Nuffield Physics Project (1957) and CERN's *Science for All* program began to align educational aims with the experimental and cooperative ethos of modern science.

The following period falls from the 1960s and the 1970s. This period could be named as "expansion of educational reform and teacher training". That is because the 1960s and 1970s consolidated these reformist efforts. In the United States, the Commission on College Physics (1960) and the National Research Council recommendations (1972) extended innovation from schools to universities, promoting inquiry-based instruction and improved teacher education. The decade also witnessed the birth of large-scale evaluation projects and diagnostic testing that began to treat learning as an object of systematic investigation.

In Europe, the diffusion of Nuffield materials and the adoption of project-based learning spread more gradually but helped shift attention toward experimentation and the process of scientific inquiry. Teacher training programs across several countries began integrating laboratory-based learning and didactic research. However, traditional systems of examination and academic culture continued to limit structural change.

Then we move to the period around the 1980s which could be remembered as "the birth of Physics Education Research and conceptual learning". Essentially, The 1980s marked a turning point as Physics Education Research (PER) emerged as a self-aware academic field. Scholars such as Arons, Karplus, and McDermott developed methodologies for probing student reasoning and documented

systematic misconceptions. The creation of the Force Concept Inventory (Halloun and Hestenes, 1985) was emblematic of this empirical turn, introducing tools for quantitative assessment of conceptual understanding.

At the same time, computational technologies began transforming laboratory practice. Early computer-based simulations, data acquisition systems, and modeling tools reshaped experimentation and visualization in classrooms. In Europe, the Institute of Physics spearheaded reforms in physics teaching and higher education, including the introduction of integrated MPhys/MSci degrees in the UK, reflecting a professionalization of physics training.

Now we shift to the 1990s, a period characterised by institutionalization and technological integration. In fact, by the 1990s, PER had achieved recognition as a legitimate research domain. Studies such as Hake (1998) demonstrated the measurable effectiveness of active learning and interactive engagement strategies, shifting attention from traditional lecture-based pedagogy to student-centered approaches. In this same decade, conceptual physics flourished, making physics more accessible to broader audiences, while computational physics solidified its role as a third pillar of scientific inquiry alongside theory and experiment.

In Europe, initiatives like the UK's *16-19 Physics Project* (1997) aimed to modernize curricula and counter declining enrollments. Across both continents, digital tools were increasingly incorporated into teaching, allowing students to visualize systems, simulate experiments, and model physical phenomena dynamically.

The 2000s–2010s can be named the period of "research-based practice and laboratory reform". In that period the consolidation of PER brought renewed focus on laboratories, representation, and student agency. Influential position papers and reports from the AAPT (1998; 2015) emphasized the importance of hands-on exploration, conceptual reasoning, and collaboration. Laboratory courses increasingly integrated digital interfaces and data logging, reflecting a broader epistemic shift from procedural accuracy toward cognitive engagement.

During the same period, European physics education became more research-oriented, with many countries developing their own communities of didactics and educational research. Inquiry-based science education (IBSE) gained institutional support through EU-funded projects, contributing to a shared vocabulary of *inquiry*, *representation*, and *modeling*.

Finally, from the 2010s to the present, we face the most recent phase that reveals a diversification of epistemic interests. Physics education now encompasses research on socio-scientific issues, environmental sustainability, gender and inclusion, and more recently, the implications of artificial intelligence for teaching and learning. The concept of representational competence has become a cornerstone, linking cognitive, linguistic, and semiotic dimensions of learning. Computational methods and data-rich analysis are increasingly present, both as research tools and as topics of investigation.

Concluding, this historical synthesis was indispensable for interpreting the model's output. Awareness of these events and changing of the educational interests and policies allowed us to contextualize clusters corresponding to distinct eras or pedagogical paradigms and inform the interpretation both looking at the period of publications and the type of content that potentially was influenced by any

specific interest of the time. The reconstruction thus functions as a methodological safeguard against anachronistic readings of the model, aligning computational findings with disciplinary history.

Engaging with this history exemplifies the human–machine complementarity that is a key pillar of computational methods for textual data analysis presented in the previous section. While the LDA model identifies latent structures through statistical regularities, it is historical and theoretical knowledge that transforms these distributions into meaningful narratives. The interpretive reconstruction anchors the computational analysis within the epistemic tradition of physics education, showing that modelling and meaning-making are inseparable activities in the analysis of scientific discourse.

2.4.7 Results

The inductive analysis of the complete corpus (1966-2019) revealed a rich landscape of twenty topics that characterise the literature published in *PE* and *TPT*. Together, these topics capture both the disciplinary and pedagogical dimensions of physics education, encompassing content knowledge, teaching practice, and learning research. The complete list is presented in Table 2.2, which reports each topic's label and top-weighted words. For interpretive clarity, topics are grouped as follows: content-based topics (1-14, blue), corresponding to the main domains of physics (e.g. mechanics, waves, optics, electricity and magnetism, thermodynamics, astronomy, and modern physics); Topic 15, in green, which focuses on the history and philosophy of physics, bridging disciplinary knowledge with its cultural and epistemological contexts; teaching-focused topics (16-19, pink), addressing course structures, curriculum design, laboratory instruction, demonstrations, and data acquisition; Finally, Topic 20, in light brown, that captures learning-focused approaches, marking the intersection between classroom practice and insights from PER.

Each topic was carefully examined to better understand how the teaching and learning of physics evolved across the period analysed. For these topics, the interpretation involved connecting the weighted words identified by the model with representative articles, allowing a contextual reading of their thematic significance.

For example, in the case of Topic 15, History and Philosophy of Physics, it was observed that this topic brings together articles exploring the lives and contributions of key scientists (e.g., *"Albert Einstein: His Life"*, *Physics Education*, 1979), the evolution of scientific thought (*"Assimilation and Development in Arab Scientific Thought"*, *The Physics Teacher*, 1966), and the role of historical texts (*"Early Science Books and Their Women Translators"*, *The Physics Teacher*, 1998). It also includes works that discuss the intersection of physics with broader cultural and philosophical dimensions, such as *"A Non-believer Looks at Physics"* (*Physics Education*, 1987). Collectively, it shows how historical narratives and philosophical reflections have been used to shape and renew our understanding of physics across time.

A different example is topic 17, about programs, textbooks, and teaching strategies, which appears predominantly in *TPT*. This topic spans curriculum design, textbook selection, teaching strategies, and assessment practices. Representative articles include *"Student Selection of the Textbook for an Introductory Physics Course"* (2007) and *"How Do You Select Your Physics Textbook?"* (1989), both addressing the interaction between instructors and students in choosing instructional materials, thereby reflecting ongoing debates about the pedagogical role of textbooks. Other contributions, such as *"Update on the Status of the One-Year, Non-Calculus Physics Course"* (1994) and *"Retests: A*

Better Method of Test Corrections" (2011), illustrate how the journals have served as platforms for discussing course structures and alternative assessment strategies aimed at improving student learning through feedback and iteration.

Table 2.2 - Topics discovered through the inductive thematic analysis of complete dataset (articles from TPT and PE journals), with corresponding top words. Content-based topics are highlighted in blue (Topics 1-14), the topic about history and philosophy of physics in green (Topic 15), teaching-focused topics in pink (Topics 16-19), and the learning-focused topic in light brown (Topic 20).

Topic number	Topic name	Top-weighted words per topic (LDA model)
Topic 1	Fundamental equations and mathematical principles in physics	2,9%"equation" + 1,0%"constant" + 0,8%"unit" + 0,8%"mass" + 0,7%"term" + 0,6%"equal"
Topic 2	Newtonian mechanics	5,7%"force" + 2,4%"motion" + 1,9%"acceleration" + 1,9%"mass" + 1,8%"object" + 1,7%"body"
Topic 3	Applied mechanics	3,7%"ball" + 2,2%"speed" + 1,5%"velocity" + 1,1%"collision" + 1,0%"fall" + 0,9%"motion"
Topic 4	Fluid mechanics and gas laws	4,6%"water" + 2,5%"pressure" + 2,2%"air" + 1,8%"liquid" + 1,4%"volume" + 1,1%"tube"
Topic 5	Waves and sound	4,7%"frequency" + 3,6%"wave" + 3,1%"sound" + 1,4%"signal" + 1,0%"amplitude" + 0,8%"pulse"
Topic 6	Colour and light	4,9%"light" + 1,8%"image" + 1,5%"colour" + 1,5%"color" + 1,2%"spectrum" + 1,1%"camera"
Topic 7	Optics	2,6%"light" + 2,6%"image" + 2,1%"lens" + 1,5%"mirror" + 1,3%"object" + 1,3%"ray"
Topic 8	Electromagnetism and currents	4,2%"magnet" + 3,0%"magnetic_field" + 2,9%"field" + 2,9%"coil" + 2,4%"current" + 2,3%"magnetic"
Topic 9	Electrostatics, and atmospheric physics	2,2%"charge" + 1,6%"surface" +0,9%"plate" + 0,7%"wind" + 0,7%"air" + 0,7%"area"
Topic 10	Circuits and electronics	2,4%"circuit" + 2,3%"voltage" + 2,0%"current" +1,3%"resistance" + 0,8%"output" + 0,7%"power"
Topic 11	Thermodynamics	5,8%"temperature" + 4,5%"energy" + 2,4%"heat" + 1,1%"power" + 1,0%"cool" + 0,8%"water"
Topic 12	Astronomy	2,7%"earth" + 1,9%"sun" + 1,7%"star" + 1,1%"astronomy" + 1,0%"planet" + 1,0%"space"
Topic 13	Nuclear physics and medical physics	1,0%"radiation" + 0,8%"sample" + 0,7%"electron" + 0,7%"energy" + 0,7%"cell" + 0,7%"detector"
Topic 14	Fundamental particles and interactions	4,2%"particle" + 3,3%"energy" + 2,2%"electron" + 1,3%"mass" + 1,3%"quantum" + 1,1%"atom"
Topic 15	History and philosophy of physics	0,7%"book" + 0,5%"world" + 0,5%"scientist" + 0,5%"theory" + 0,5%"scientific" + 0,4%"idea"
Topic 16	Trends in physics education: course structures and career pathways	0,8%"pupil" + 0,7%"subject" + 0,6%"education" + 0,6%"project" + 0,4%"examination" + 0,4%"unit"

Topic 17	Programs, texts, and teaching strategies	1,1%“class” + 1,0%“program” + 1,0%“high_school” + 0,7%“college” + 0,7%“lab” + 0,7%“text”
Topic 18	Demonstrations and apparatus in physics teaching	3,0%“fig” + 1,2%“apparatus” + 0,8%“length” + 0,8%“demonstration” + 0,8%“cm” + 0,7%“spring”
Topic 19	Data acquisition and measurement in physics education	2,5%“data” + 1,7%“measurement” + 1,1%“graph” + 1,0%“sensor” + 0,9%“model” + 0,7%“experimental”
Topic 20	Innovative learning strategies in physics education	1,8%“learn” + 1,0%“activity” + 1,0%“question” + 1,0%“concept” + 0,9%“group” + 0,9%“research”

Major trends and temporal evolution

The temporal analysis of the topics distribution revealed broad yet coherent trajectories in how physics education discourse evolved between 1966 and 2019. These trajectories reflect both the institutional history of the journals and the epistemic transformations of the field itself. Figure 2.3 illustrates the percentage prominence of each topic over time, with colours corresponding to the four groupings to support a synoptic reading of the field’s evolution.

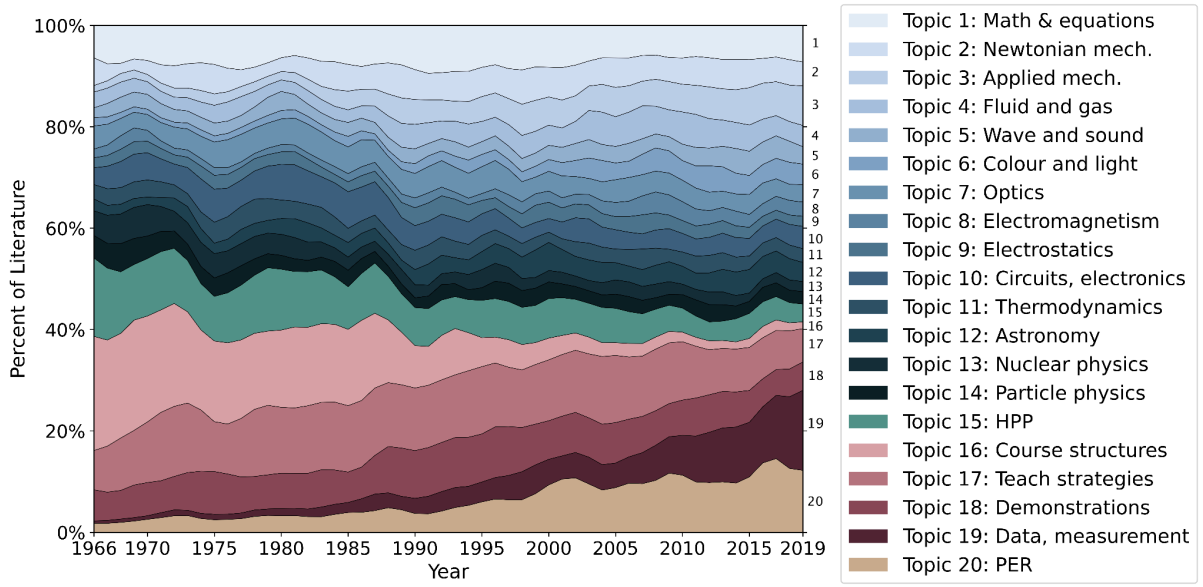

Figure 2.3 - Stacked area plot displaying the normalized prevalence of 20 topics in *Physics Education* and *The Physics Teacher* from 1966 to 2019 (with a three-years rolling window). The plot illustrates the percentage prominence of each topic over time, with content-based topics (blue) at the top, followed by History and philosophy of physics (green), teaching-focused topics (pink), and the PER-focused topic (brown) at the bottom.

Some key trends:

- **Early expansion and the consolidation of teaching discourse (1966-1973)**: In the initial period, the journals experienced a rapid increase in publication volume, reflecting their growing recognition within the physics education community. The thematic focus was primarily on *teaching practice*: demonstrations, classroom experiments, and the articulation of physics principles for effective instruction. This period established the journals as

platforms devoted to improving pedagogy through accessible, experience-based teaching approaches.

- **Hands-on learning, reform, and the rise of research-based teaching (1980s-1990s):** The late 1980s marked a crucial turning point. Topics related to *demonstrations and apparatus* (Topic 18) surged, signalling a collective effort to bridge disciplinary content with active learning experiences. The historical context of educational reform and the emergence of PER are reflected in the model's structure: alongside apparatus and laboratory innovation, a new topic began to emerge (Topic 20), focused on *learning processes and evidence-based teaching*. This period corresponds to the institutional birth of PER as a research area concerned with conceptual understanding, student misconceptions, and inquiry-based methods. The appearance of topics linking *teaching practice* with *learning research* suggests that the field was beginning to internalise a new epistemology: that improving teaching requires understanding how students learn.

- **Technological integration and the expansion of PER (2000–2020):** From the early 2000s onward, the model captures a decisive shift toward *computational and data-driven approaches* in physics education (Topic 19). Articles increasingly discuss microcontrollers, sensors, and simulations as tools for experimentation and modelling. This reflects a broader digital transformation in education, as well as a new conception of experimentation itself, mediated by computational representations. In parallel, topics associated with *learning strategies and PER findings* continued to grow, indicating the consolidation of PER as a mature field. The founding of *Physical Review Physics Education Research* (PRPER) in 2005 marked this institutional shift, legitimising PER within the physics community. The growing prominence of these topics in *The Physics Teacher* and *Physics Education* illustrates how research-informed teaching became an integral part of mainstream practice.

The inductive thematic analyses conducted with the support of the LDA model, both on combined corpus of TPT and PE together and on the journals separately, allowed us to identify interesting trends and features. Below, we highlight the most significant.

The enduring prominence of physics content is identified. Despite these transformations, the analysis also revealed remarkable continuity. Content-based topics (1-14) remained dominant across the decades, representing roughly half of the corpus throughout the period. This is clearly visible in Figure 2.3, where the portion of content-based topics (blue topics) constitutes more or less 50% of the topics of the two magazines constantly over time. This persistence demonstrates that the discussion of physics content has remained central to the identity of both journals, serving as a bridge between disciplinary expertise and pedagogical reflection. Moreover, the strong overlap between content and pedagogical topics (articles that blend explanations of physical phenomena with strategies for teaching them) suggests that content has become increasingly "pedagogised": it is discussed not for its own sake, but as a vehicle for understanding how students engage with concepts.

Then a significant shift from teaching to learning is recognised. When content-oriented topics are excluded, like in Figure 2.4, the remaining distribution (Topics 16-20) reveals a striking shift from teaching-focused to learning-focused discourse. In the earlier decades, the literature concentrated on course structures, career pathways, and teaching strategies, essentially on the organisation of instruction. From the mid-1990s onward, attention turned toward understanding student learning, conceptual change, and research-based instructional design. This shift aligns with the increasing

influence of PER and cognitive science, marking a broader methodological transition in physics education from prescription to inquiry.

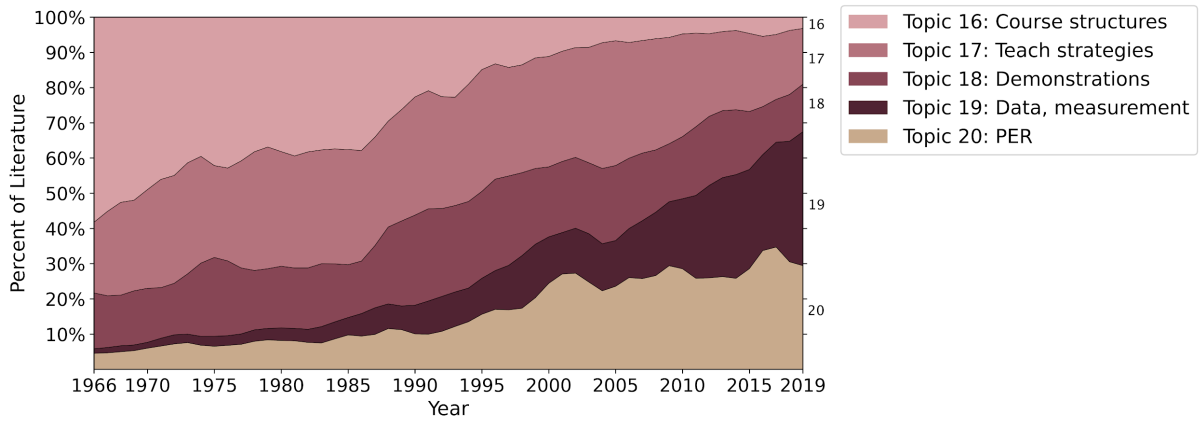

Figure 2.4 - Stacked area plot showing the normalized prevalence of topics 16 to 20 in *Physics Education* and *The Physics Teacher* from 1966 to 2019 (with a three-years rolling window). This plot illustrates the percentage prominence of each topic over time, highlighting the shift from teaching-focused topics (16-18, pink) to measurement and data (topic 19, purple) and learning-focused topic (topic 20, brown).

The transformation is both quantitative and qualitative: the number of publications on Topics 19 and 20 rises sharply, while the conceptual centre of gravity moves from describing how teachers teach to analysing how learners construct meaning. This evolution reflects a profound epistemic reorientation, from knowledge transmission to knowledge co-construction.

By comparing the two journals, distinct yet complementary profiles were recognised. *TPT* has historically maintained a balance between content and pedagogy, characterised by stability and an enduring emphasis on the history, philosophy, and practicalities of teaching. Over time, it progressively incorporated technological innovation and PER findings, reflecting the American community's responsiveness to research-based reform movements. This is recognisable in Figure 2.5, that shows the staked area plot of normalised prevalence, which is the percentage of each topic appearing over time.

In contrast, *Physics Education* shows a more dynamic evolution, visible also in Figure 2.6., which in the same way as figure 2.5, takes a "time-lapse" of the PE articles. Initially centred on curriculum design and educational pathways, it gradually shifted toward content, pedagogy, and PER, mirroring European efforts to harmonise educational practices across diverse national systems. Its earlier engagement with PER in the 1990s indicates a proactive adoption of research-informed methodologies, while *The Physics Teacher* followed this trend more strongly in the 2000s. Together, the journals illustrate two complementary traditions: one pedagogically stable and historically grounded, the other research-oriented and progressively adaptive.

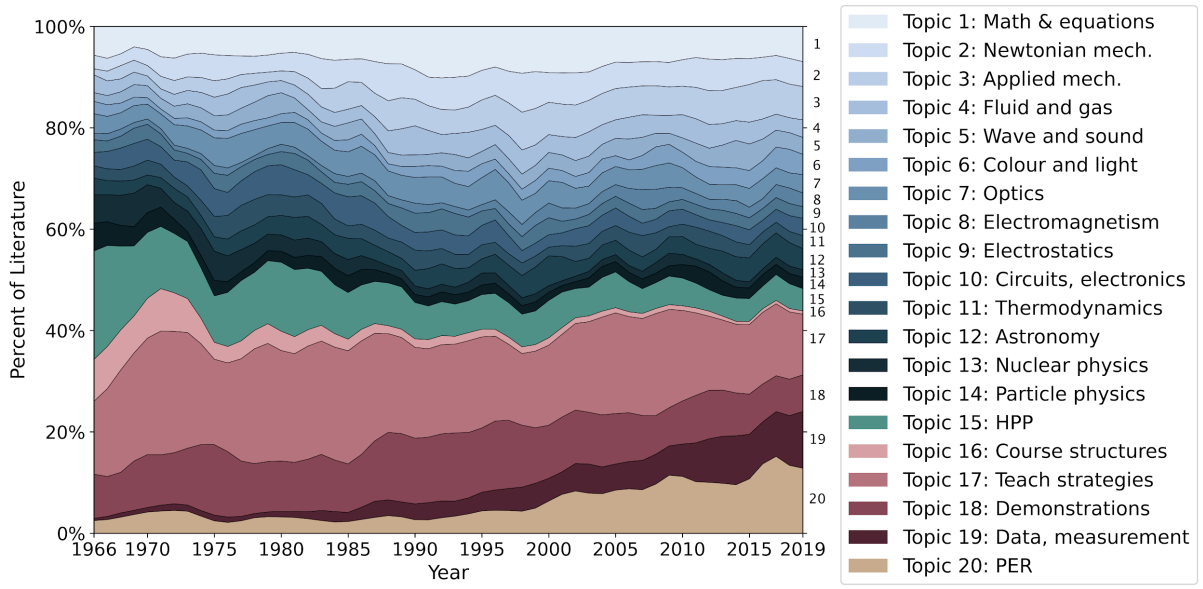

Figure 2.5 - Stacked area plot displaying the normalized prevalence of 20 topics in *The Physics Teacher* from 1966 to 2019 (with a three-years rolling window). The plot illustrates the percentage prominence of each topic over time, with content-based topics (blue) at the top, followed by History and philosophy of physics (green), teaching-focused topics (pink), and the PER-focused topic (brown) at the bottom.

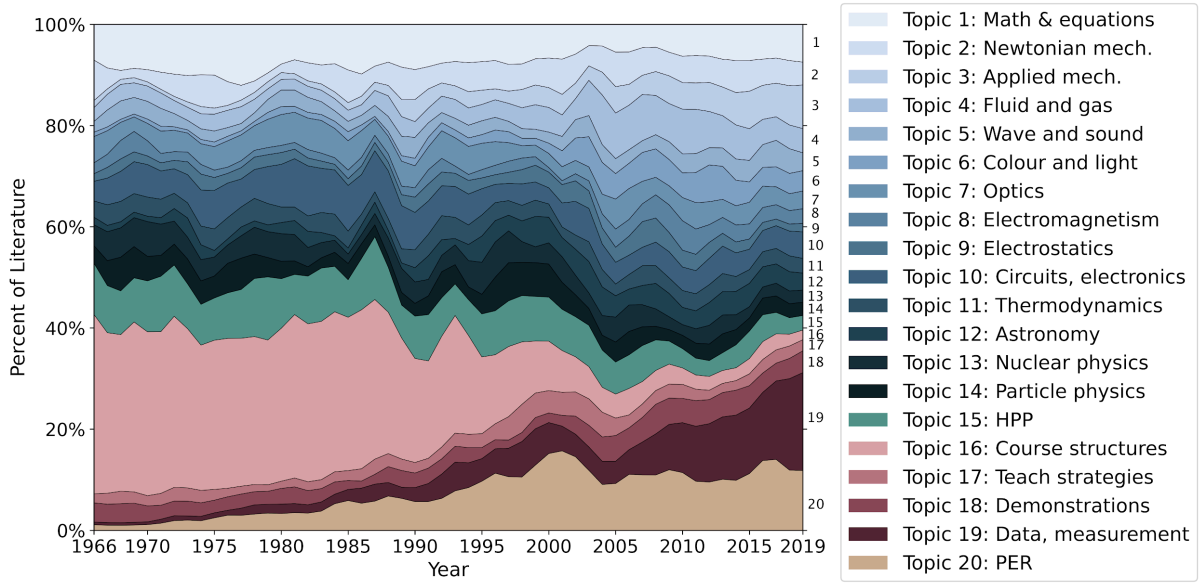

Figure 2.6 - Stacked area plot displaying the normalized prevalence of 20 topics in *Physics Education* from 1966 to 2019 (with a three-years rolling window). The plot illustrates the percentage prominence of each topic over time, with content-based topics (blue) at the top, followed by History and philosophy of physics (green), teaching-focused topics (pink), and the PER-focused topic (brown) at the bottom.

2.4.8 Discussion

Beyond its empirical outcomes, this case study contributes to a broader understanding of how physics education has evolved as a field of knowledge production. The topic trends identified through LDA

can be read not only as patterns in language but also as traces of successive epistemic regimes within science education.

Historically, the analysis mirrors the waves of educational reform that shaped physics teaching during the post-war period. The early expansion of *TPT* and *PE* in the 1960s and 1970s coincided with international initiatives to modernise science education amid the Cold War, as nations invested in scientific literacy and curriculum renewal. Both journals served as institutional mediators, translating these reforms into professional and pedagogical discourse. Over time, their content documents a gradual shift from the didactic transmission of physics knowledge toward reflective, inquiry-based, and research-informed teaching practices.

The results also highlight two forces that have acted as epistemic accelerators within the field. The first is technological innovation, which redefined experimentation and data acquisition, from manual demonstrations to computational and sensor-based learning environments. The second is the consolidation of PER, which introduced systematic, evidence-based methodologies and shifted attention from teaching strategies to students' conceptual understanding and cognitive processes. Together, these drivers transformed not only how physics is taught, but also how knowledge about teaching is generated and validated.

Several limitations must be acknowledged. First, the dataset reflects published discourse rather than observed classroom practice. The journals analysed offer insight into how educators and researchers *write about* physics teaching, but not necessarily how teaching is enacted in practice.

Second, the analysis is inherently shaped by the characteristics of the LDA model and its bag-of-words representation. While LDA offers transparency and interpretive accessibility, it identifies statistically coherent clusters of words rather than conceptually defined topics. Consequently, the resulting themes represent dominant linguistic regularities (many of them content-based) rather than a comprehensive taxonomy of pedagogical concerns.

Third, the research design required deliberate curatorial choices to balance algorithmic constraints with educational relevance. Vocabulary selection, topic number, and parameter tuning were adjusted to ensure that key educational terms (e.g., "learn", "question", "activity") were retained and that non-content topics could emerge. These interpretive interventions were essential for meaningful results but also constitute potential sources of subjectivity.

Fourth, the study is limited to two English-language journals, representing primarily Anglo-American and European perspectives. Although their long publication history and international readership make them valuable indicators of broader trends, they cannot capture the full diversity of global physics education traditions.

Finally, the analysis does not include the most recent transformations in the field, such as the post-pandemic digital turn or the integration of artificial intelligence in education. These developments fall beyond the temporal scope of the dataset and remain important directions for future inquiry.

Looking ahead, newer natural language processing and AI-based approaches offer promising continuations and extensions of this work. Text embedding methods, for instance, provide richer semantic representations by capturing contextual relationships between words, while LLMs, as discussed by Wulff, Kubsch, and Krist (2025), enable more complex analyses of educational texts,

from thematic synthesis to discourse interpretation. Moreover, integrative frameworks such as computational grounded theory (Nelson, 2020) and thematic network analysis (Mariegaard et al., 2022) demonstrate how human interpretive reasoning and computational scalability can be combined for transparent and reproducible text analysis. These developments point toward the future of mixed-methods inquiry in Science Education Research, where computational models serve not only as analytical tools but also as epistemic partners in reflective methodology.

However, this case study illustrates how a relatively simple model such as LDA can enable a methodologically aware analysis. Precisely because of its transparency and well-understood mechanics, every modelling decision (e.g. vocabulary selection, parameter tuning, or topic interpretation) was open to scrutiny and reflection. Rather than treating the algorithm as a black box, I could observe how large textual data are transformed, reduced, and reorganised through computational operations, and how these transformations invite human interpretation. The process thus became an opportunity to learn how human and algorithmic reasoning can cooperate in large-scale inductive analyses, revealing how meaning can be progressively constructed through the interplay of model design, data structure, and researcher judgement.

2.4.9 Concluding reflections for case study A

The identification of meaningful topics in this study required sustained interpretive work at multiple levels: the careful curation of the dataset, the iterative preparation and tuning of the model, and the contextual interpretation of its outcomes. The reconstruction of the historical and institutional evolution of physics education was essential to inform the analysis, providing the background against which the model's statistical regularities could acquire meaning.

A crucial aspect of the process was the continuous triangulation between different sources of knowledge: the model outputs, the historical reconstruction, and our own disciplinary and methodological expertise. My co-supervisor, professor Odden, and I repeatedly engaged in a reflective dialogue, comparing the model's constraints with contextual information and theoretical sensitivity. We soon realised that none of these elements, algorithmic, historical, or interpretive, was sufficient in isolation. Only by orchestrating and harmonising them could coherence be achieved.

In different moments of the process, one component would temporarily take precedence over the others. For instance, when choosing the optimal LDA configuration, we first relied on a quantitative criterion, called the elbow method, to identify a plausible range of topic numbers. Yet, the final decision was made through researcher judgement, grounded in our understanding of the historical evolution of the field. This interplay between computational precision and human sensitivity exemplifies the kind of methodological hybridity that this thesis seeks to illuminate.

Although the process can be described through distinct phases (such as data preparation, topic development, and topic interpretation), these stages were not linear. They were dynamically connected in an iterative cycle, where insights from one phase constantly reshaped the next. Stability and meaning emerged only through this recurrent dialogue between model and researcher.

On a personal level, this experience profoundly shaped my understanding of what it means to engage in data-intensive research in science education. The model does not "speak" by itself: to interpret its signals, I had to immerse myself in the historical and socio-cultural contexts of physics education, exploring how each period reflected specific educational values and societal priorities. What might

appear as a limitation of the model, its inability to generate meaning autonomously, became instead a source of epistemic enrichment, compelling me to integrate contextual and theoretical knowledge with computational evidence.

Conversely, the use of a computational model opened possibilities that would be unattainable by manual analysis. It enabled the examination of more than thirteen thousand articles, making visible statistical regularities and long-term dynamics that no human researcher could detect unaided. This capacity to see the field from both macro and micro perspectives, tracing large-scale trends without losing the individual voice of single publications, epitomises the hybrid methodology that underpins this thesis.

Looking back, what remains most striking is the sense of continuity that emerges from this long temporal view. Reading titles and abstracts written decades ago, I encountered the same educational intentions that guide physics teaching today: the desire to engage students, to make abstract concepts tangible, and to cultivate curiosity about the physical world. Despite changes in tools, technologies, and theoretical frameworks, the core pedagogical impulse remains the same. Recognising this enduring thread while analysing it through contemporary computational means has been both methodologically instructive and personally moving. This experience thus bridged past and present through the lens of a model that, while artificial, ultimately helped to reveal something profoundly human about the pursuit of teaching and learning physics.

2.5 Second case study: thematic analysis and topic modeling for analysing students' essays

2.5.1 Rationale and context

This second case study remains situated within the research field of science education but explores a different situation in terms of research focus, dataset type, and scale from the previous one. While the first case involved a large corpus of over twelve thousand research articles and employed *Latent Dirichlet Allocation* to map trends in the literature, the present study concentrates on a medium-sized, context-specific textual dataset of about 200 students' essays written in Italian language. The data volume is smaller yet still intensive: large enough to challenge manual analysis, but too limited to qualify as "big data". Therefore, it creates a boundary condition neither easily affordable through qualitative nor data-intensive approaches. Certainly, a methodological tension of possibilities arises. On one hand, there are qualitative methods, which in SER are traditionally employed to deal with textual data, valued for their capacity to preserve contextual meaning and interpret students' voices. On the other hand, AI techniques offer researchers new computational possibilities. As digital infrastructures increasingly mediate educational activity, methods such as text mining and topic modelling provide alternative ways to approach textual data.

For this reason, such a situation becomes an ideal terrain for methodological reflection. The kinds of questions it raises are familiar to many SER researchers:

- How can textual data be analysed while preserving the nuance of context-specific language?
- What scale of data can be realistically managed through human-driven analysis?
- Which methods allow for knowledge generation that remains faithful to the research question and science education context and needs?

These questions are not merely technical but epistemological. They concern the type of knowledge that textual data can offer, how analytical methods shape and transform that knowledge, and what is gained or lost in the process.

It is within this landscape that the present case study is positioned, where I and several colleagues explored possible ways to analyse a dataset produced within a research project on future-oriented science education. The corpus was composed of essays written by high-school students who were asked to describe their lives in the year 2040. Analysing these essays presented a dual challenge: on the one hand, the texts were rich and contextually grounded, making them ideal for qualitative interpretation; on the other hand, their number and linguistic variability made a fully manual analysis difficult.

Consequently, this case study considers two different methods, qualitative and computational, as possible candidates for analysing the same corpus, with the dual purpose of exploring textual data through different approaches and comparing how they produce and interpret knowledge. The aim is to move beyond procedural comparison and to reflect on the relation between method, data, analysers and knowledge, that are central concerns of this thesis.

This section draws on and expands the chapter "Comparing Textual Data Analysis Methods in Science Education Research" (Caramaschi, Zanellati & Levrini, 2025) published in the book of selected proceedings of ESERA 2023. Figures and tables are adapted from that publication.

2.5.2 Aims, research questions, and methodological design

The study was designed as a methodological exploration situated at the intersection of qualitative and computational paradigms in Science Education Research. Its central aim was to examine how different analytical approaches shape the kind of knowledge that can be generated from the same textual dataset. Specifically, it compares a *Reflexive Thematic Analysis* (TA), representative of the qualitative, interpretive tradition, and a *Latent Semantic Analysis* (LSA), representative of data-intensive, computational approaches.

The specific purpose was twofold. First, to investigate how each method processes meaning and abstraction when applied to medium-sized, context-specific textual data, that in this case is a corpus of 223 essays written by high-school students about their imagined futures in 2040. Second, to analyse what epistemological features and methodological choices underpin each approach, and how these influence interpretation, generalisation, and validity in the resulting analyses.

This design deliberately isolates the analytical process as the object of inquiry. Both methods were applied to the *same dataset*, under the same research aim, allowing for a direct comparison of how qualitative and computational logics differently construct and constrain meaning. The research questions guiding this case study were therefore methodological and epistemological:

1. How does the choice between a qualitative and a computational data-driven method influence the kind of knowledge produced when analysing medium-sized, context-specific textual data in Science Education Research?
2. What intrinsic features of each method - epistemological and procedural - enhance or limit their interpretive potential?

In this case study, both analyses were carried out within the same research group but through different configurations of collaboration. I participated directly in both. For the Thematic Analysis (TA), I led the data analysis together with a group of colleagues, developing a comprehensive qualitative study that has since been published; this process is described in detail in Section 2.5.4.

For the Latent Semantic Analysis (LSA), the computational work was primarily conducted by my colleague Andrea Zanellati. My role was to support the interpretation phase by contextualising the data and clarifying the educational aims of the corpus.

Then, for the purpose of this thesis, we together reconstructed the entire LSA analytical process (see Section 2.5.5), enabling me to examine it in parallel with the TA process. Based on these two reconstructed analyses, I subsequently developed two analytical grids designed to investigate and compare their epistemological and methodological features. These grids constitute the framework for the comparative analysis presented in the final part of this case study in Section 2.5.6.

2.5.3 Dataset and context of data collection

The dataset analysed in this case study was produced within the FEDORA project (*Future-Oriented Science Education to Enhance Responsibility and Agency*[5]), funded by the European Union's Horizon 2020 programme (2020–2023). The project aimed to explore how science education can help young people imagine and engage with possible futures, developing scientific literacy that is both epistemically grounded and socially responsive. Within this framework, the University of Bologna research group in Physics Education focused on investigating students' representations of the future as a way to understand how they make sense of science, society, and their own agency within both.

The data consisted of 223 essays written by Italian high-school students between 2018 and 2021. These essays were collected during extracurricular courses organised through the *Piano Lauree Scientifiche* (PLS) initiative, which is an Italian national programme promoting scientific engagement among secondary school students. The participating schools were located in different regions of Italy and represented a range of socio-economic contexts. Students took part voluntarily as part of enrichment activities linked to themes such as *climate change*, *simulation and computing*, *artificial intelligence*, *seismic risk*, and *quantum technologies*.

Each student was invited to respond to the same open-ended writing prompt, designed by the research group to stimulate creative but reflective reasoning about the future. The task asked students to imagine their lives in the year 2040, and to describe:

1. where they imagine living and what kind of life they lead;
2. the problems or challenges they and their communities face;
3. the opportunities, technologies, and environments that surround them; and
4. the kinds of social relations and forms of participation they envision.

This task structure ensured a shared thematic frame across the corpus while leaving students ample freedom to express personal ideas, emotions, and reasoning. As a result, the texts varied in style and depth. For example, some did a brief descriptive narrative imagining a day in their future life, while

[5] FEDORA is an EU-funded project on future-oriented Science Education to enhance responsibility and engagement in the society of acceleration and uncertainty. Web site: https://www.fedora-project.eu/

others decided to write longer, reflective essays. Most ranged between 300 and 800 words, creating a dataset that is moderately extensive in size and highly heterogeneous in content.

From a methodological perspective, this corpus offers an ideal testing ground for exploring analytical approaches. The essays are context-rich, embedded in a clear educational activity, and written in natural language that reflects authentic student expression rather than elicited responses to questionnaires. At the same time, the medium volume of data presents a significant analytical challenge: it exceeds what can be exhaustively examined through traditional qualitative methods, yet it is not large enough to justify fully data-intensive modelling.

An additional layer of complexity arises from the Italian language of the texts. Working in a non-English corpus required specific preprocessing and linguistic adaptation when applying computational techniques such as Latent Semantic Analysis. Moreover, the imaginative and speculative nature of the writing (e.g. students reasoning about hypothetical futures) posed interpretive challenges that required both sensitivity to context and methodological flexibility.

These characteristics make the dataset a fertile ground for methodological exploration. It represents a case of textual data situated in a *liminal zone* between qualitative and quantitative conditions, strongly reminding us that, regardless of the data's nature, the crucial difference lies in *how* one chooses to approach the analysis and *what* methodological decisions are made along the way.

2.5.4 Thematic Analysis (qualitative approach)

Method overview

The qualitative analysis followed the principles of Thematic Analysis (TA), which is a qualitative approach widely adopted across disciplines such as psychology and the social sciences (Braun & Clarke, 2006; 2019). Rather than a single, unified technique, it represents an umbrella term encompassing diverse approaches that share the goal of identifying and interpreting themes within qualitative data. Themes, understood as patterns of shared meaning, capture elements that are particularly relevant to the research question and illuminate significant aspects of the phenomenon under investigation (Braun & Clarke, 2006).

Although TA does not prescribe a rigid sequence of steps, Braun and Clarke propose a set of guiding phases that can support the analytical process. The first involves familiarisation with the data, through close and repeated reading to gain an in-depth understanding of the material. This is followed by coding, in which short labels are assigned to segments of text that appear meaningful or potentially relevant to the research questions. The next phase entails generating initial themes by grouping related codes, and subsequently reviewing and refining these themes to ensure coherence and distinction. The final phase consists of defining, naming, and interpreting the themes, and producing an analytical narrative that accounts for the entire process.

A key strength of TA lies in its theoretical flexibility. It is not bound to a particular epistemological or methodological position, and can thus be adapted to different research traditions and objectives (Braun & Clarke, 2019). This flexibility, however, requires methodological rigour to ensure credibility and transparency. Braun and Clarke highlight four principles that help maintain this balance:

1. **Clarity of guiding questions**, since themes must respond to the specific aims of the study;

2. **Sensitivity of the researcher**, as thematic discovery relies on interpretive awareness rather than mechanical procedures;
3. **Trustworthiness of the process**, achieved through clear and transparent analytic practices such as triangulation; and
4. **Groundedness in the dataset**, meaning that themes should be meaningful within the data context rather than determined by frequency alone.

Description of the analysis using RTA

The analysis of the students' essays is described in detail by Barelli and colleagues (2022). In this section, only the main stages of the analytical process are summarised, following the scheme in Figure 2.7.

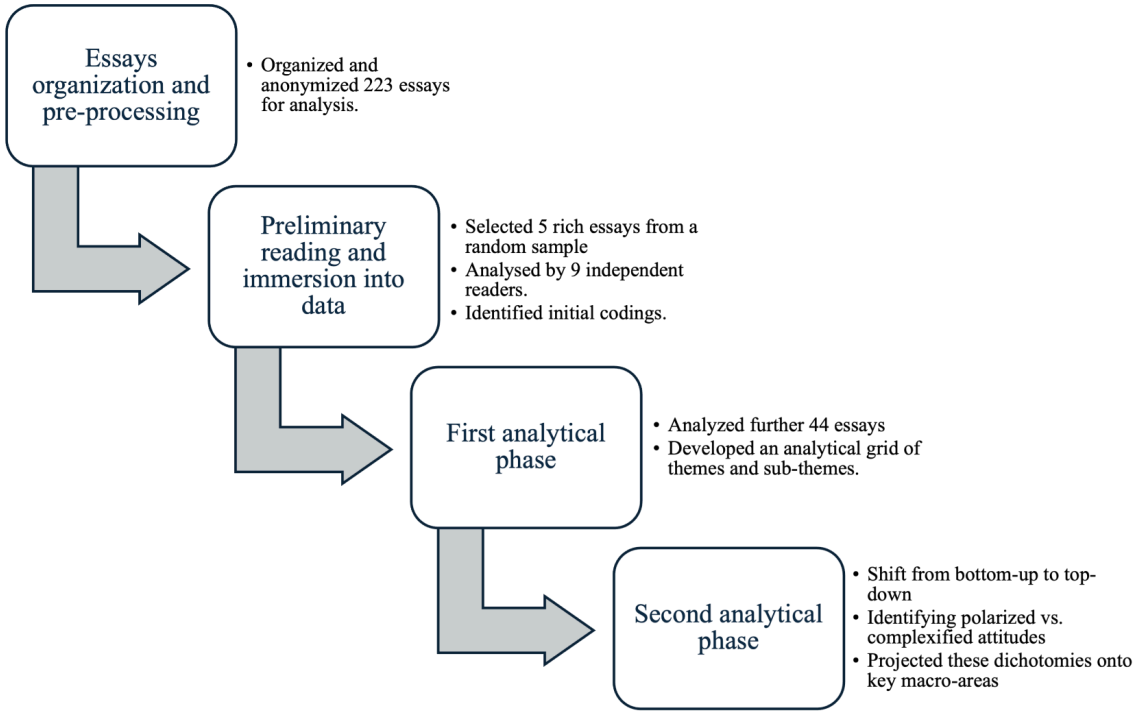

Figure 2.7 - The RTA was implemented by performing four main phases: Essays organization and pre-processing, preliminary reading and immersion into data, a first analytical phase, and a second analytical phase.

Students' essays on the future were examined through "reflexive thematic analysis" (Braun & Clarke, 2019) which is one specific type of TA that considers the researcher's subjectivity and reflexivity as a central feature. The analysis was conducted with a "bottom-up" strategy aimed at allowing patterns to emerge directly from the data rather than imposing predefined categories.

After an initial close reading of five essays, the research team, composed of nine members, independently produced preliminary codes and subsequently collaborated to construct the first set of thematic categories. The analysis was then continued by four researchers, who examined an additional forty-four essays through an iterative process that alternated between phases of individual immersion in the data and collective triangulation. During this first analytical phase, we identified nine overarching themes and a series of sub-themes that represented more specific manifestations of each main theme (see Figure 2.8, from Barelli et al., 2022). The team deliberately maintained a bottom-up approach, remaining as closely aligned as possible with the data. However, as the analysis progressed,

we collectively recognised that this inductive process was not converging toward saturation. Instead, it was generating an increasingly detailed and extensive thematic map that risked replicating the data rather than abstracting meaning from it.

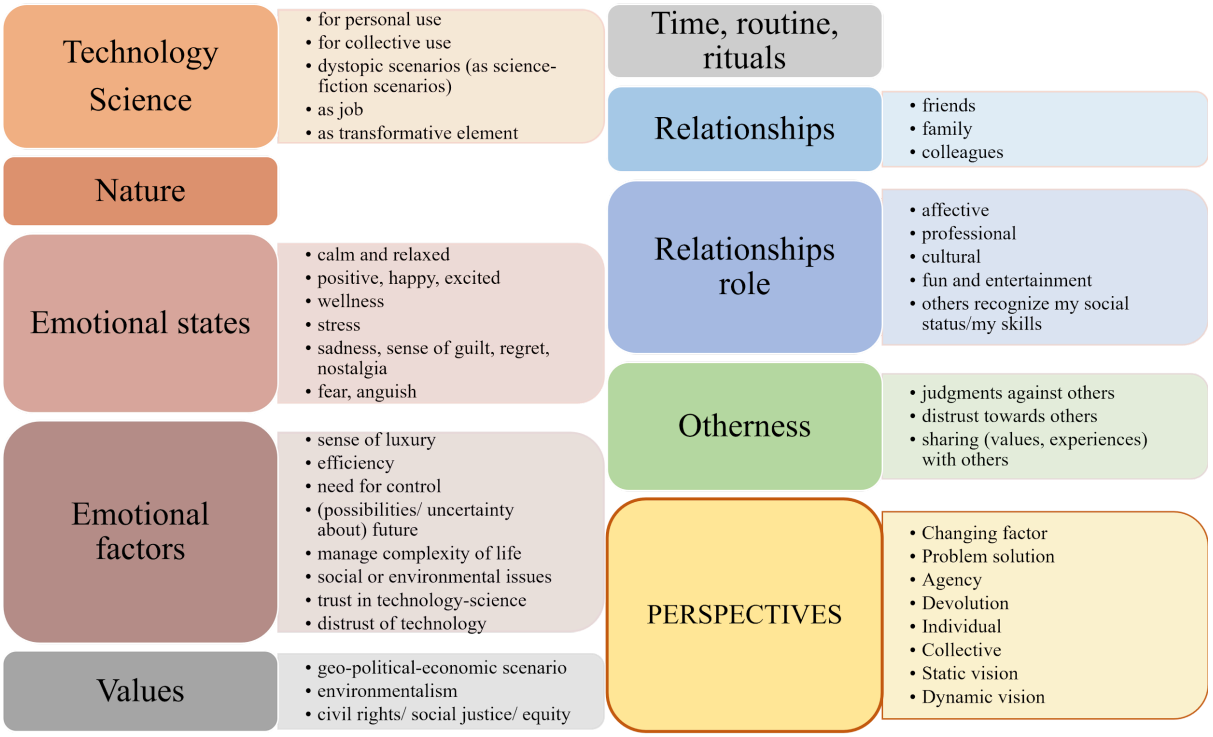

Figure 2.8 - A visualization of the grid's macro-themes, sub-themes, and perspectives from the first analytical phase.

Another element that was not working was that, after immersing themselves in the data, the researchers recognised that some essays reflected a *polarisation attitude*. This concept, previously discussed as a critical issue in science education studies, refers to a tendency to interpret complex issues in dichotomous terms, reducing them to two opposing and mutually exclusive positions. In such cases, nuance and ambiguity are simplified, leading to rigid judgements that prioritise taking a stance over fostering understanding. We expected that the coding could capture this aspect of the essays, but it did not. Maintaining a purely bottom-up stance generated an increasingly detailed structure that risked mirroring the data rather than interpreting it. The thematic map was rich in description but lacked integrative explanatory power. Reflexive memos and peer debriefings led the team to recognise that saturation had not been reached at a conceptual level: the analysis was expanding horizontally instead of deepening. Also, the inductive approach therefore failed to adequately represent the theoretical construct of *polarisation attitude*. Recognising this impasse prompted a deliberate shift in our analytic stance. In the second phase, we stepped back from the data, by then well familiar, and reoriented our attention toward the underlying phenomenon perceived within it, engaging in a form of axial coding (Vollstedt & Rezat, 2019).

The team decided to shift focus from *what* topics students mentioned to *how* they constructed their futures—what stance, tone, and structure characterised their narratives. Returning to the data with this axial lens revealed two contrasting attitudes of imagining:

1. Complexification, where students described futures acknowledging trade-offs, interdependencies, and uncertainties;

2. Polarisation, where futures were portrayed in dichotomous or deterministic terms, such as utopia versus dystopia or progress versus catastrophe.

This second phase also resulted in the development of an analytical grid (see Table 2.3 adapted from Barelli et al., 2022) that proved capable of generating new insights. The grid offered a more coherent interpretive framework through which the two contrasting attitudes could be meaningfully distinguished. This was a turning point reaching an achievement that had eluded the initial inductive analysis.

Table 2.3 - Grid developed in the second analytical phase, resulted from a set of dichotomies (columns) projected onto the macro-areas (rows).

		Dichotomies				
		Personal-Societal	Functional-Aesthetics oriented	Good-Bad	Natural-Artificial	Certain-Uncertain
Macro-areas	Science and technology	Personal use of S&T - Societal use of S&T	Humans rule on machines - Machines rule on humans	Negative - Positive impact of S&T	Nature as the priority - Technology as the priority	Certain - Uncertain of S&T scenarios
	Relationships	Self-centred circle of family and friends - Societally-oriented relationships	Useful for wellbeing - Generative of delight	Extreme harmony - Difficulty of cultivating relationship	Authentic face-to-face relationships - Relationships mediated by technology	Fixed and established - Undefined or in continuous transformation
	Emotions	Self-determined - Societally-determined emotions	Determined by a functional-oriented life - Determined by aesthetics-oriented life	Pessimism - Optimism	Visceral - Artificially induced emotions	Need for control - Refusal of control
	Time and space routines	Self-determined - Societally-determined routines	Determined by the search of performance - Determined by the search of delight	Routine increases the quality of life - routines are oppressive	Routine immersed in nature - Routine dominated by technology	Fixed routines - Vague or absent routines
	Occupation	Individual career - Career with social impact	Performance - Aspiration	Good role in someone's life - Bad or absent role in someone's life	In contact with nature - Totally immersed in technology	Already decided career - Unimagined career

Throughout the process, reflexivity played a central role. Researchers kept analytic diaries to note interpretive shifts, discussed alternative readings in regular meetings, and re-examined raw data whenever a theme or category was revised. Rather than eliminating subjectivity, these practices made it visible and accountable, supporting credibility through transparency. Triangulation among analysts,

prolonged engagement with the corpus, and explicit reflection on theoretical sensitivity all contributed to the trustworthiness of results.

The RTA ultimately produced two complementary outcomes. First, a descriptive thematic framework representing the main areas of meaning in students' future narratives; and second, a conceptual lens, about the complexification/polarisation distinction, captured in the analysis grid. More than a set of results, this process exemplified how methodological insight arises through the negotiation between bottom-up immersion and top-down theorisation. The analysis demonstrated that generating educational knowledge from textual data requires not only sensitivity to participants' voices but also the courage to reshape interpretation when inductive exploration reaches its limits.

2.5.5 Latent Semantic Analysis (computational approach)

Method overview

The second analysis applied a Latent Semantic Analysis (LSA) to the same corpus of students' essays. LSA is a computational technique from the family of vector-space models, rather than probabilistic topic models (Landauer et al., 1998; Dumais, 2004). Within the field of Education Data Science, it is valued for its transparency and interpretability in uncovering the semantic structure of textual data (themes). The method applies Singular Value Decomposition (SVD), a form of principal component analysis (PCA), to a document-term matrix, reducing its dimensionality and revealing patterns of word co-occurrence. In this way, LSA identifies latent conceptual relationships among words and documents, highlighting how terms cluster together across the corpus to form underlying semantic themes. Statistical weighting enters during the pre-processing stage, where documents are represented as TF–IDF matrices that balance word frequency within and across texts (see details in the next paragraph).

While more recent deep-learning models (e.g., transformer-based LLMs) learn language representations from vast datasets, LSA relies on a simpler and more mathematically transparent principle: the assumption that the meaning of a word can be approximated by the company it keeps. By reducing the complexity of natural language into a multidimensional semantic space, the algorithm identifies clusters of terms that tend to occur together, inferring the hidden topics that could plausibly have generated the observed word patterns.

Reconstruction of the analytical workflow using LSA

In this project, the LSA analysis was carried out by my colleague Andrea Zanellati, with my involvement focusing on the interpretive and contextual phases. My contribution was to clarify the educational objectives of the corpus, support the pre-processing of Italian-language data, and later help reconstruct the analytical workflow for the purpose of the present methodological comparison.

The analysis was implemented in Python within the *Google Colaboratory* environment, using common machine-learning libraries such as *pandas* and *scikit-learn.* Pre-processing represented a crucial phase because the corpus was in Italian and composed of students' creative writing, which is semantically rich but syntactically irregular. A custom text-cleaning and lemmatization pipeline was therefore developed. The steps included:

- Stopword removal, using the *Natural Language Toolkit (NLTK)* Italian stopword list, extended with domain-specific words (e.g., “student”, “essay”, “teacher”) that were contextually frequent but semantically uninformative.
- Lemmatization and tokenization, performed using *TreeTagger*, with a manually refined dictionary to merge words sharing the same lemma but not automatically recognised.
- Vocabulary refinement, producing a final lexicon of approximately 1,500 unique terms, sufficiently compact to capture recurrent semantic relations while preserving the dataset’s thematic diversity.

After pre-processing, each essay was represented as a vector of weighted terms using the TF–IDF transformation (Term Frequency–Inverse Document Frequency). This step balanced the contribution of frequent and rare words, ensuring that terms unique to specific essays received higher weight than very common ones.

The next stage consisted of constructing the term–document matrix, which was then decomposed using Singular Value Decomposition (SVD), the mathematical core of LSA. SVD projects the high-dimensional textual data into a reduced semantic space, allowing words and documents to be compared based on cosine similarity. In this space, proximity indicates semantic relatedness. Clusters of words forming coherent regions correspond to latent *topics* or *conceptual areas*.

A critical analytical choice concerned the number of dimensions, that is, how many latent topics would meaningfully describe the corpus. Following Dumais (2004), this number was determined empirically through model testing rather than predetermined theoretically. Several candidate models were computed and visually inspected for interpretability, with five main topics ultimately retained.

Interpreting the topics required moving from mathematical abstraction back to educational meaning. The model produced lists of the most strongly weighted words per topic, which then had to be semantically labelled. Initially, this was attempted in a bottom-up fashion by examining each list of words and seeking a unifying concept that could capture its shared meaning. However, the resulting groupings were often too generic (e.g., topics dominated by words like *life, people, world, work*), yielding clusters that lacked specificity and interpretive value (see Figure 2.9).

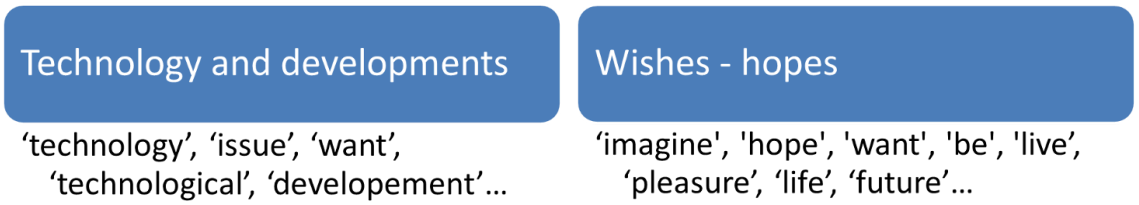

Figure 2.9 - Examples of lists of words (translated from italian) for two different topics identified in the analysis of LSA

To improve interpretability, we adopted a contextually informed, top-down procedure. External knowledge about the social and temporal context of the data was reintroduced as a lens for interpretation. The essays had been written between 2018 and 2021, a period marked by major societal events likely to shape students’ imaginaries, most notably, the global climate protests initiated by Greta Thunberg in late 2018 and the COVID-19 pandemic beginning in 2020. We therefore reorganised the corpus into three temporal subgroups (2018, 2019–2020, 2021) and examined how the distribution of topics changed across these periods.

To visualise and quantify patterns, we applied a K-means clustering algorithm to group documents in the latent space. The overlap between clusters and year groups was evaluated using similarity measures such as *f1-score*, *intersection-over-union*, and *cosine similarity*. This multi-step inspection supported the interpretive phase by identifying which clusters were temporally or semantically coherent.

The interpretive synthesis resulted in five topics, each described through its ten most representative words. Despite the clarity of the computational pipeline, interpretation remained difficult. Some topics appeared too broad or trivial (e.g., "people, world, life"), while others overlapped semantically, reflecting the limited size and linguistic homogeneity of the corpus. The medium data volume and heavy lemmatization, although necessary for model stability, had reduced textual nuance, constraining the model's ability to distinguish subtle variations in meaning.

From a methodological standpoint, the LSA analysis highlighted both the power and fragility of computational approaches to text in educational contexts. Its power lies in systematically mapping the semantic space of a corpus that would be too large for exhaustive human reading, revealing global structures and recurrent lexical associations. Yet its fragility lies on how much it depends on pre-processing and parameter choices: every decision, from the removal of stopwords to the number of dimensions retained, affects what patterns the model can "see".

In the context of this study, these findings underscored that computational methods, while mathematically elegant, still require researcher judgement at multiple stages. Interpretation cannot be automated; rather, it demands theoretical awareness of the dataset's provenance, language, and purpose. The analysis also illuminated how such models privilege regularities and frequency, potentially overlooking the idiosyncratic or affective dimensions of students' expression, elements that qualitative approaches are better equipped to capture.

The reconstruction of this LSA workflow was therefore essential for the comparative aim of the study. It made visible the chain of methodological assumptions embedded in the computational process and provided the basis for analysing its epistemological features alongside those of the Reflexive Thematic Analysis. In both cases, meaning emerged not from the data alone but from a dialogue between algorithmic modelling and researcher interpretation, exemplifying the hybrid character of methodological practice at the intersection of AI and qualitative inquiry.

2.5.6 Comparison of analytical processes and results

The comparison between the RTA and the LSA was conducted using two analytical grids developed specifically for this case study. The first grid examined the *epistemological quality* of the knowledge generated by each method, focusing on three key dimensions: generalisability, interpretability, and originality. By *generalisability*, we refer to the degree to which the analytical outcomes provide insights that can be validly extended to the dataset as a whole. This notion resonates with the concept of *saturation* as discussed in Grounded Theory (Vollstedt & Rezat, 2019), where analytical stability indicates that new data would not substantially alter the emerging understanding. *Interpretability* relates to the production of themes or topics that are both representative and meaningful, achieved through a process that balances attention to detail with a comprehensive appreciation of the contextual nuances of the data (Kristanto & Padmi, 2020). Finally, *originality* captures each method's potential to generate new knowledge or, at least, to uncover fresh and previously unexpressed perspectives on the phenomenon under study.

The second grid analysed the *methods feature*s, comparing them in terms of the resources and conditions underpinning the analytical process. These included the types and amount of human and technical resources required, the specific skills and methodological expertise involved, and the strategies adopted to engage with and manipulate the data. Furthermore, this tool examined the particular conditions under which each method was applied, thereby highlighting their respective strengths, limitations, and domains of applicability.

Together, these two grids provided a tool to systematically contrast the two analytical approaches, not only in their outcomes but also in the underlying logics through which they produce novel knowledge. Table 2.4 reports the results of the comparison between RTA and LSA in relation to methods features, while Table 2.5 presents the corresponding results for the epistemological quality of the two methods.

Table 2.4 - Grid for methods comparison on three methods features (taken from Caramaschi, et al., 2025).

	Methods features		
	Analytic resources	Approach to data	Applicability conditions
Thematic Analysis Bottom-up phase	The analytical process involved nine researchers. Several iterations of analysis, triangulation, and refinement were necessary.	Researchers' engagement with data, and sensitivity to capture student voices. Spiral iterative process: a few essays at a time were analyzed to develop the macro and sub-themes. Overall, 49 essays out of 223 were analysed.	When saturation is not reached, the method remains limited not only to the context but also to the subset of the dataset. Disproportion between the human resources available for the analysis and dataset dimension.
Thematic Analysis top-down phase	Theoretical insight is used to look at data, enabling the development of a new analytical lens. A second analytical phase is performed.	Researchers detach from data and then come back to look at them, through the sensitising idea that emerged from the insight.	Since at this point essays had already been read, and the analysis grid was able to synthesize the phenomenon, its application was fast.
Latent Semantic Analysis	The analytical process involved only one researcher, who had computational and science education expertise. It required a scrupulous and customized preprocessing phase.	Iterative refinement of topics, automatically retrieved, improving, in turn, the list of removed and aggregated words. All essays were analysed.	Scalable method to reuse in the future, if additional data will be collected. Issue given by a slight difference in the essay task over years. Borderline quantity of texts for a data-driven approach.

Table 2.5 - Grid for methods comparison on three epistemic quality descriptors (taken from Caramaschi, et al, 2025).

	Epistemic quality		
	Generalisability	Interpretability	Originality
Thematic Analysis Bottom-up phase	Macro themes appeared to be generalisable, in agreement with the literature. Instead, sub-themes identification did not saturate, turning into a one-to-one description of the essays, without grasping a generalizable knowledge among all essays.	The deep immersion in the data allowed the analysers to recognise clear themes, reaching a good level of interpretability of macro and sub-themes.	The granularity of sub-themes was too fine and widespread, preventing the detection of high-level patterns or phenomena in the data. Original results were found when we detached from TA, reformulating the sub-themes as dichotomies.
Thematic Analysis top-down phase	The axial coding that led to the development of the analysis grid used to identify "complexified" and "polarized" attitudes, allowed us to grasp the general phenomenon emerging from data. Saturation was reached.	Themes and dichotomies that composed the analysis grid were highly interpretable thanks to a deep understanding of data combined with literature knowledge.	The analysis grid, thanks to internal coherence and operational capacity, leads to identifying patterns in data. Thus, it can be considered an instrument generator of knowledge.
Latent Semantic Analysis	The LSA model is built on the whole corpus of essays. Thus, every essay can be described as a combination of the discovered topics. New essays are likely to be described by the same topics.	Often a topic was described by generic words. Thus, meaning inferring was trivial. Moreover, an overlapping of semantics between the retrieved clusters of words hindered the interpretability of results.	The scarcity in results granularity and topic vagueness made them not very informative. Causes could be identified in the medium-sized corpus and reduction of word complexity, which limited the power of the model.

The development of these grids will be presented in the next chapter. Here, we focus on the analytical results obtained through their application in this case study. Specifically, we provide an overview of the two methods' features, describing the resources required for analysis, the researchers' approaches to data, and the conditions under which each method was applied, as well as how knowledge was generated and which epistemic moments fostered the creation of original, generalisable, and interpretable results.

The most striking difference between the two analyses lies in the position of the analyser in relation to the data. In TA, knowledge emerged through *prolonged human engagement* (e.g. through cycles of reading, coding, and reinterpretation) where meanings were constructed through dialogue within the research team. Reflexivity and collaboration were epistemic engines of the analysis: every interpretive move was situated, debated, and recorded through memos and meetings. The process was therefore hermeneutic and iterative, oriented toward saturation and conceptual coherence. By contrast, in LSA, topic development and interpretation were distributed mainly between the algorithm and the researcher. The algorithm performed the primary abstraction, translating language into vectors and co-occurrence patterns beyond the reach of manual human analysis. The researcher's role shifted from immersion to design and calibration: deciding how to preprocess text, set model parameters, and interpret output. This delegation of pattern discovery to a computational model introduced a layer of mediation that enabled systematic analysis of all essays but reduced direct phenomenological contact with the data. Moreover, the fragmentation of information, from texts to words, made it difficult to reconstruct meaning a posteriori.

Another key difference concerns time and scale. TA required sustained involvement from nine researchers over several months of iterative analysis, producing depth and contextual sensitivity but limited coverage (49 of 223 essays). LSA, conducted by one researcher in a shorter period, encompassed the entire corpus, achieving breadth and consistency at the expense of nuance. In this sense, TA and LSA embody opposite epistemic strategies: the former privileges intensive interpretation of fewer cases; the latter, extensive mapping of larger datasets.

A third distinction lies in the nature of abstraction. TA builds abstraction through interpretive synthesis, linking codes into themes and reorienting them toward higher-order categories (e.g., complexification/polarisation). Each abstraction remains traceable to concrete excerpts. LSA, instead, constructs abstraction mathematically, compressing the semantic space into latent dimensions not directly anchored to textual instances. While this allows for generalisation and replicability, it detaches meaning from its original narrative context. The two methods thus model meaning differently: one through contextual condensation, the other through statistical projection.

Regarding the analytical results, the two analyses generated different epistemic qualities. The RTA generated two levels of findings: a thematic map that captured the main content areas of students' future narratives, and a conceptual framework distinguishing *complexification* and *polarisation,* explaining variations in how students imagined the future. This explained differences in how students imagined the future. These results were interpretive, grounded in close reading, and embedded in theoretical sensitivity. They provided rich, nuanced insights but were limited in scope due to the partial subset of essays analysed in depth.

The LSA, on the other hand, yielded five computationally derived topics, each described through lists of words with varying interpretability. These results provided a panoramic but flattened view of the corpus, identifying recurrent lexical associations rather than conceptual structures. The topics revealed broad tendencies (e.g., frequent references to work, environment, or technology) but lacked the capacity to articulate the attitudes or relationships among these ideas. The findings were thus synthetic but thin: they illuminated the semantic terrain without penetrating its qualitative depth.

From an epistemological standpoint, two main insights emerged. First, the analyses revealed complementary strengths and limitations, as highlighted in the literature. TA demonstrates interpretive richness and contextual anchoring but faces scalability constraints and potential overfitting to

analysts' perspectives. LSA offers scalability and consistency but filters out contextual and affective dimensions central to educational meaning. This supports the idea that computational and qualitative analyses have potential for integration within mixed-methods frameworks. However, overcoming each method's limitations requires more than applying both: it demands the thoughtful design of mixed-method approaches that genuinely integrate their complementary features.

The second insight concerns where original knowledge is generated. In both analyses, the greatest epistemic value lies in the interplay between bottom-up and top-down processes, where new understanding is effectively produced. The bottom-up process enables the exploration of data and the identification of potential patterns, while the top-down, theory-driven interpretation provides coherence and meaning. Although not a novel observation, the close comparison of TA and LSA makes this epistemic interplay particularly evident. This interplay is the locus where the new Generative AI can provide further insights and contributions, questioning and exploring the nature of who or what contributes to the generation of knowledge.

2.5.7 Reflections and conclusions

The confrontation between the Reflexive Thematic Analysis and the Latent Semantic Analysis provides more than a methodological comparison; it offers a window into the evolving methodological landscape of SER. What emerges is not a hierarchy between qualitative and computational approaches, but a recognition of their complementarity and the distinct epistemic functions each fulfils within the study of educational phenomena.

The two methods embody different ways of approaching textual data and, more profoundly, different ways of conceiving what it means to "know" through text. TA constructs meaning through interpretive engagement, in which researchers immerse themselves in the data, iteratively reading, coding, and theorising in relation to the context. It thrives on proximity and dialogue between analyst and data, making it particularly sensitive to the nuances, metaphors, and contradictions of students' imaginative writing. LSA, by contrast, models meaning through distance and generalisation. It operates by translating language into distributions and co-occurrences, making visible regularities that transcend individual expression. Its strength lies in its ability to extend analysis to larger corpora, enabling the exploration of broad semantic landscapes that would otherwise be inaccessible. Yet, it is precisely in this abstraction that meaning risks losing its connection with context, emotion, and voice.

This difference, however, should not be read as incompatibility. Rather, it reflects a productive methodological tension. The rich, context-bound and interpretive nature of educational data requires approaches that oscillate between *immersion* and *abstraction*, *closeness* and *distance*. TA and LSA thus form two poles of a continuum of inquiry, each compensating for the other's limits. The former provides depth and contextual grounding; the latter provides breadth and systematic coverage. Neither is sufficient alone to grasp the full epistemic complexity of text-based data in education.

One insight that emerges strongly from this comparison is that both methods require interpretation and theoretical anchoring. Computational models such as LSA do not "speak" on their own: their output gains meaning only when interpreted through educational theory and contextual knowledge. Likewise, qualitative themes are not self-evident discoveries but analytic constructions shaped by the researcher's perspective and the conceptual frameworks they mobilise. In both cases, *theoretical transparency* and *reflexivity* are indispensable to ensure that results remain scientifically meaningful and not merely descriptive or technical.

Another observation concerns scale and applicability. The present study has shown that LSA, although applicable to medium-sized corpora, unfolds its full potential in larger-scale analyses, where the volume of text supports the statistical assumptions underlying topic modelling. In contrast, when dealing with highly contextualised, creative, or linguistically rich texts, such as students' narratives about the future, the complexity and polysemy of language challenge the model's capacity to generate distinct and interpretable topics. This calls for careful adaptation and hybridisation of computational tools when applied to educational data.

In summary, the comparison confirms that themes and topics, though superficially similar as semantic units, are epistemically distinct. Themes arise from interpretive synthesis; topics from statistical structure. Yet both can converge toward understanding when interpreted in relation to context and theory. The point is not to choose between them, but to articulate their dialogue within a coherent methodological framework.

This reflection directly informs the heuristic tool developed in Chapter 3. The contrasts and complementarities identified here constitute its conceptual foundations. The tool builds on the idea that methods are not neutral instruments but epistemic models, each embodying specific assumptions about what counts as data, pattern, and knowledge. By making these assumptions explicit, the heuristic aims to guide researchers in recognising the methodological, epistemological, and axiological dimensions of their analytical choices.

Ultimately, this case study demonstrates that the integration of qualitative and computational methods does not simply multiply techniques; it transforms the very paradigm of inquiry in Science Education Research. As researchers engage with both human and machine-based analyses, they are called to develop new forms of methodological literacy, capable of navigating between interpretive understanding and computational modelling, and of articulating how each contributes to the shared endeavour of making sense of educational meaning in the age of data and AI.

2.6 Pillars characterising computational AI-based methods in SER

This chapter has moved between theory and practice, combining a conceptual exploration of computational methods with two empirical case studies that made these reflections tangible. The first case study examined large-scale literature through topic modelling, while the second investigated students' essays through a dual analysis using Thematic Analysis and Latent Semantic Analysis. Across these experiences, methodological reflection became inseparable from practice: the act of applying, comparing, and interpreting methods exposed their epistemic logics and limits more vividly than theoretical discussion alone.

Taken together, these explorations revealed that computational approaches to textual data in Science Education Research are not merely technical procedures but *epistemic acts*: ways of modelling, reducing, and transforming meaning. When viewed through this lens, the qualitative and computational traditions, rather than opposing one another, appear as complementary articulations of a shared methodological space. From this realisation, several pillars emerge that underpin the practice of computational and hybrid inquiry.

Pillar 1: Human-Machine complementarity

A first foundational insight concerns the interdependence of human interpretation and algorithmic modelling. In both case studies, meaningful understanding arose only when the researcher actively engaged with computational outputs, naming topics, contextualising patterns, and relating them to educational significance. The algorithm alone could not produce interpretive knowledge, just as human reading alone could not capture the statistical structures revealed by the model. This complementarity reframes the researcher's role: rather than a user of tools, the researcher becomes a mediator between analytical logics, navigating the translation between statistical abstraction and contextual meaning.

Pillar 2: Text representation and reduction

Every computational method involves transforming language into another form of representation, probabilistic, vectorial, or network-based. This transformation necessarily entails a reduction of linguistic complexity, a selective compression that enables analysis while constraining interpretation. The first case study, for instance, reduced educational discourse to word distributions, while the second simplified Italian narratives into lemmatized tokens. Recognising these reductions is not a limitation but an epistemic responsibility: what is gained in tractability is balanced by what is lost in nuance. Thus, representation and reduction must be treated not as neutral operations but as methodological choices that determine what kind of knowledge a model can produce.

Pillar 3: Relationality of knowledge

A third shared pillar is the relational conception of knowledge that underlies both human and computational inquiry. In topic models, meaning arises from patterns of co-occurrence (relations among words and documents) while in thematic analysis, meaning emerges from connections drawn among codes, themes, and contexts. In both cases, understanding is constructed through relations rather than isolated elements. This relationality aligns computational epistemology with broader notions of complexity and systems thinking in science education, suggesting a common ground between data-driven pattern recognition and interpretive synthesis.

Pillar 4: Contextual grounding and situated meaning

Both approaches demonstrated that analytical methods, regardless of their degree of automation, require contextual anchoring. In the TA, the meaning of themes was inseparable from students' cultural and educational experiences; in the LSA, topic interpretation only became coherent when informed by contextual knowledge, such as temporal events like climate protests or the pandemic. Context operates as a semantic field that sustains interpretation: without it, both algorithms and analysts risk generating decontextualised or misleading generalisations. Methodological awareness in SER therefore requires maintaining a continuous dialogue between data, context, and theory.

Pillar 5: Benefit of epistemic transparency and reflexivity

Finally, both experiences underscored the need for epistemic transparency, a clear articulation of the assumptions, constraints, and interpretive moves that shape the research process. In qualitative analysis, transparency was achieved through reflexive documentation of coding decisions and interpretive shifts. In computational analysis, it required explicit reporting of preprocessing steps, model parameters, and interpretive criteria. Transparency thus becomes a shared ethical and

methodological value: it enables reproducibility in computational terms and trustworthiness in interpretive terms. Reflexivity, in turn, ensures that methodological choices are not concealed behind technical authority but are recognised as constitutive of the knowledge produced.

References

American Association of Physics Teachers; Goals of the Introductory Physics Laboratory. *Am. J. Phys.* 1 June 1998; 66 (6): 483–485. https://doi.org/10.1119/1.19042

American Association of Physics Teachers. 2015 Annual Report. https://aapt.org/Publications/upload/2015_Annual_Report.pdf

Badilescu, S. (1998) Early Science Books and Their Women Translators. Phys. Teach. 36, 516–518 https://doi.org/10.1119/1.880123

Barelli, E., Tasquier, G., Caramaschi, M., Satanassi, S., Fantini, P., Branchetti, L., & Levrini, O. (2022). Making sense of youth futures narratives: Recognition of emerging tensions in students' imagination of the future. Frontiers in Education, 7, 1-17. https://doi.org/10.3389/feduc.2022.911052

Blei, D. (2011). Introduction to Probabilistic Topic Models. *Communications of the ACM, 55*, 77-84. https://doi.org/10.1145/2133806.2133826

Braun, V. & Clarke, V. (2006) Using thematic analysis in psychology. Qualitative Research in Psychology, 3:2, 77-101. https://doi.org/10.1191/1478088706qp063oa

Braun, V. & Clarke, V. (2019). Reflecting on reflexive thematic analysis. Qualitative Research in Sport, Exercise and Health, 11:4, 589-597. https://doi.org/10.1080/2159676X.2019.1628806

Dumais, S.T. (2004). Latent Semantic Analysis. Annual Review of Information Science and Technology (ARIST), 38, 189-230.

Caramaschi, M., & Odden, T. O. B. (2025). Analyzing the history of physics education in the USA and Europe through natural language processing. Physical Review Physics Education Research, 21, 020153. https://doi.org/10.1103/wfvw-hkyy

Caramaschi, M., Odden, T. O. B., & Levrini, O. (accepted, in press). Reviewing 55 years of Literature on Physics Education using Natural Language Processing.

Caramaschi, M., Zanellati, A., & Levrini, O. (2015). Comparing Textual Data Analysis Methods in Science Education Research. Chapter in "Connecting Science Education with Cultural Heritage" Springer book. DOI: 10.1007/978-3-031-90944-3

Dake, L.S. (2007). Student Selection of the Textbook for an Introductory Physics Course" The Physics Teacher, 45, 7, 416–419. DOI: 10.1119/1.2783148

Ganascia, J. G. (2010). Epistemology of AI Revisited in the Light of the Philosophy of Information. *Knowledge, Technology & Policy,* 23, 57–73. https://doi.org/10.1007/s12130-010-9101-0

George, S. (1994) Update on the status of the one-year, non-calculus physics course. The Physics Teacher, 32, 5, 344–346. DOI: 10.1119/1.2343097

Hake, R. (1998). Interactive-Engagement Versus Traditional Methods: A Six-Thousand-Student Survey of Mechanics Test Data for Introductory Physics Courses. American Journal of Physics - AMER J PHYS. 66. 10.1119/1.18809.

Hake, E. J. (2011). 'Retests': A better method of test corrections. The Physics Teacher, 49, 3, 168–171. DOI: 10.1119/1.3555393

Halloun, I. A., & Hestenes, D. (1985). The initial knowledge state of college physics students. Am. J. Phys. 53 (11): 1043–1055. https://doi.org/10.1119/1.14030

Hermann, B. (1987). A non-believer looks at physics. Phys. Educ. 22 280 DOI 10.1088/0031-9120/22/5/316

Humphreys, P. (2004). Extending Ourselves: Computational Science, Empiricism, and Scientific Method. New York, US: Oxford University Press.

Kalai, A. T., Nachum, O., Vempala, S. S., & Zhang, E. (2025). Why language models hallucinate. https://doi.org/10.48550/arXiv.2509.04664

Krist, C., Kubsch, M., Wulff, P. (2025). Human-Machine Interactions in Machine Learning Modeling: The Role of Theory. In: Wulff, P., Kubsch, M., Krist, C. (eds) Applying Machine Learning in Science Education Research. Springer Texts in Education. Springer, Cham. https://doi.org/10.1007/978-3-031-74227-9_8

Kristanto, Y. D., & Padmi, R. S. (2020). Using Network Analysis for Rapid, Transparent, and Rigorous Thematic Analysis: A Case Study of Online Distance Learning. Jurnal Penelitian dan Evaluasi Pendidikan, 24(2), 177-189.

Landauer, T. Foltz, P.W., & Laham, D. (1998). Introduction to Latent Semantic Analysis. Discourse Processes, 25, 259–284. https://doi.org/10.1080/01638539809545028

Lewis, J. (1999). Chapter 9. Physics Education in “125 years: the Physical Society and the Institute of Physics” ed. by John L. Lewis, Institute of Physics Publishing, XIII, 243 p. : ill.; ISBN 0-7503-0609-2

Mariegaard, S., Seidelin, L. D. & Bruun, J. (2022). Identification of positions in literature using thematic network analysis: the case of early childhood inquiry-based science education, International Journal of Research & Method in Education, 45:5, 518-534. https://doi.org/10.1080/1743727X.2022.2035351

Meltzer, D. E. & Otero, V. K. (2015) A brief history of physics education in the United States. Am. J. Phys. 83, 447–458. https//doi.org/10.1119/1.4902397

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781

Nelson, L. K. (2020). Computational Grounded Theory: A Methodological Framework. Sociological Methods & Research, 49(1), 3–42. https://doi.org/10.1177/0049124117729703

Odden, T. O. B., Marin, A. & Caballero, M. D. (2020) Thematic analysis of 18 years of physics education research conference proceedings using natural language processing. Phys. Rev. Phys. Educ. Res. 16, 010142, 1-25 https//doi.org/10.1103/PhysRevPhysEducRes.16.010142

Odden, T. O. B., Marin, A., & Rudolph, J. L. (2021). How has Science Education changed over the last 100 years? An analysis using natural language processing. *Science Education*, 105, 653– 680. doi: 10.1002/sce.21623

Revere, R. B. (1966) Assimilation and Development in Arab Scientific Thought *Phys. Teach.* 4, 350–353 https://doi.org/10.1119/1.2351046

Rosser, W. G. V. (1979) *Albert Einstein: His Life Phys. Educ.* 14 220

Russell, S.J. & Norvig, P. (1995). Artificial Intelligence: A Modern Approach. Prentice-Hall, Englewood Cliffs, New Jersey.

Schultz, F. H. C. (1989) How do you select your physics textbook?. The Physics Teacher, 27, 4, 278. DOI: 10.1119/1.2342758

Tschisgale, P., Wulff, P., & Kubsch, M. (2023). Integrating artificial intelligence-based methods into qualitative research in physics education research: A case for computational grounded theory. *Physical Review Physics Education Research,* 19, 2. https://doi.org/10.1103/PhysRevPhysEducRes.19.020123

Vollstedt, M., & Rezat, S. (2019). An Introduction to Grounded Theory with a Special Focus on Axial Coding and the Coding Paradigm. In: Kaiser, G., Presmeg, N. (eds) Compendium for Early Career Researchers in Mathematics Education. ICME-13 Monographs. Springer, Cham. https://doi.org/10.1007/978-3-030-15636-7_4

Wulff, P., Kubsch, M., & Krist, C. (Eds.). (2025). Applying machine learning in science education research: When, how, and why? Springer. https://doi.org/10.1007/978-3-031-74227-9

Chapter 3 – Making methods visible

Tools for analysing the foundations of method and methodology

3.1 Introduction

After exploring different methods and gaining familiarity with computational AI-based techniques for textual data analysis, we recognized the need to take a further step toward mapping research methods and understanding their distinctive features and implications. As discussed in the previous chapters, the emergence of AI-based methods in educational research brings not only new technical possibilities but also significant epistemological and methodological challenges.

Chapter 3 focuses on the development of a heuristic tool designed to support critical reflection on research methods in science education, conceived as an evolution of the comparative grids developed in Section 2.5.6. The chapter begins by outlining the theoretical frameworks that guided the tools creation (Section 3.3). These include philosophical perspectives on methodology, such as ontological, epistemological, and axiological dimensions, alongside literature from science education that addresses the tensions and synergies between qualitative and computational approaches.

Building on these foundations, the chapter then reconstructs the process through which the tools were progressively designed (Section 3.4). This involved several iterative phases of conceptualization, testing, and refinement, informed by both theoretical insights and practical considerations. The intention was never to establish which method is more effective, but rather to offer a means of mapping, understanding and comparing different approaches. Thus, the goal was to uncover the underlying logics and assumptions of each method, thereby enhancing methodological transparency.

The outcome of this process is twofold. First, two comparative grids were developed to enable structured reflection on how methods act and operate on knowledge, encompassing dimensions such as the role of the researcher, the nature of the data, and the kinds of knowledge they manipulate and generate. Second, an evolved heuristic tool was elaborated to explore *why* methods operate as they do, allowing the examination of the foundational assumptions underlying their methodology (Section 3.5). Crucially, they serve as a reflective aid for researchers combining diverse approaches in mixed-methods designs, helping them to clarify, for example, what is being integrated and which assumptions underlie the manipulation and transformation of knowledge within the research process.

The chapter concludes by illustrating how these instruments can be applied in practice and become operational. Through these descriptions, the chapter demonstrates how the proposed instruments can foster deeper methodological awareness and support more informed and transparent research decisions.

3.2 Framing the need for reflexive tools

The increasing use of computational and AI-based techniques in science education research (SER) has introduced novel analytical possibilities while simultaneously challenging traditional methodological boundaries. As these approaches gain prominence, the need arises to understand not only how they operate technically, but also how they reshape the epistemological, ontological, and axiological

assumptions that underpin research. The introduction of AI-based methods demands an explicit reconsideration of what counts as valid knowledge, how such knowledge is generated, and what values guide its production. Yet, systematic tools for comparing and interpreting methods across such diverse traditions remain scarce.

This awareness emerged from direct engagement with different analytical practices throughout the case studies described in the previous chapter. Working with both qualitative and computational methods revealed that technical mastery alone was insufficient to fully comprehend the implications of methodological choices. In particular, while applying a thematic analysis and a topic-modelling approach to the same dataset, it became clear that the two methods, though addressing the same corpus, enacted distinct relationships between data, researcher, and meaning. Understanding these differences required moving beyond technical evaluation to a level of *methodological reflection*, a level concerned not only with *what* each method produces, but *how* and *why* it does so.

This experience prompted three central questions that have guided the development of the present tools:

- *How can we systematically compare different methods?*
  In a landscape where methods proliferate rapidly, there is a need for frameworks that enable rigorous comparison, revealing how analytical strategies differ in their assumptions, operations, and results.
- *Can we recognise similar analytical phases across different data-intensive computational analyses?*
  Despite their diversity, many AI-based methods follow analogous sequences (i.e. data preparation, transformation, modelling, interpretation) each involving distinct types of decisions and expertise. Mapping these shared phases can clarify where human and machine reasoning intersect and how they co-produce knowledge.
- *Can we develop a tool to trace and understand the roles of researchers, experts, and machines throughout the analytical process?*
  AI-based approaches redistribute agency within research, with models and algorithms performing tasks once reserved for humans. A systematic way to describe these evolving roles is crucial to interpret methodological hybridity.

At the same time, these questions resonate with broader debates currently unfolding within the SER community. Recent editorials in *Physical Review Physics Education Research* have explicitly called for a "critical examination" of both quantitative (Henderson, 2017) and qualitative methods (Henderson, 2021). These initiatives signal a shared recognition that methodological traditions, long taken for granted, require renewed scrutiny. The call is particularly urgent in an era of datafication, where digital environments generate new types of data, new analytical techniques, and consequently, new epistemic risks.

Within this evolving scenario, reflecting on methodology is not a matter of abstract philosophy but a condition for responsible and transparent research practice. Without explicit attention to the assumptions that govern methods, there is a risk that they become routine procedures, applied automatically and divorced from their theoretical justification. When methods are treated as mere tools, their underlying worldviews remain invisible (e.g. what they assume to be real, knowable, and valuable). As Ding (2019) and Corbetta (2014) warn, such opacity can lead to a superficial comprehension of methods and an impoverishment of scientific inquiry itself.

Hence, the need for a reflective tool arises from both *practical* and *foundational* considerations. Practically, researchers require instruments to navigate the growing diversity of methods and to make informed choices suited to their research goals. Foundationally, there is a moral and intellectual obligation to make visible the paradigmatic assumptions that support these choices. Reflection on methodology therefore becomes a means of reinforcing the integrity, transparency, and coherence of research.

The development of the tools presented in this chapter responds to this dual need. They are designed not to rank or prescribe methods but to foster awareness, helping researchers articulate the often implicit logics, assumptions, and values that underlie methodological practice. By offering a structured space for comparison and reflection, these instruments aim to bridge the gap between technical execution and philosophical understanding, between the explicit procedures of research and the implicit frameworks that sustain them. In this sense, the reflective tools function as a bridge between the empirical work of the previous chapter and the theoretical discussion that follows. The next section therefore outlines the theoretical and philosophical foundations that informed their design, drawing on the long-standing debates on methods and methodology, the concept of paradigms, and the triadic framework of ontology, epistemology, and axiology that anchors the heuristic tool.

3.3 Guiding frameworks and ideas

The development of the comparative grids and the heuristic tool was informed by a broad and multi-layered theoretical framework. This framework combines philosophical reflection on scientific methodology with contemporary debates in science education research and with insights emerging from practical experience. Its purpose is to provide the conceptual scaffolding for understanding *why* methodological reflection is necessary and *how* it can be systematically conducted.

3.3.1 Methods and methodology: two levels of inquiry

Before approaching the philosophical underpinnings of the tool, it is crucial to clarify a distinction often blurred in educational research, that between *methods* and *methodology*. Methods refer to the concrete investigative procedures, tools, or techniques employed for data collection and analysis within a specific study (Schwandt, 2007). Methodology, by contrast, represents a higher level of reflection: the logic, principles, and criteria that govern how knowledge is developed and validated within a discipline (Corbetta, 2014). The suffix "-logy" denotes precisely this dimension of *logos*, a reflective inquiry into the rational grounds that justify methodological choices.

This distinction becomes essential in the context of AI-based and data-intensive research. Computational methods such as topic modelling or machine learning can be executed with precision, but understanding their methodological significance requires asking what kind of knowledge they produce, what they consider valid evidence, and how they transform the relationship between researcher, data, and interpretation. Without this step of methodological reasoning, methods risk being reduced to technical recipes, detached from the assumptions that sustain their legitimacy.

3.3.2 Foundational questions and the triadic framework

Philosophical inquiry into methodology traditionally pivots around three foundational questions:

1. **Ontological:** What is the nature of the reality under investigation?
   Epistemological: What kind of knowledge can be obtained about that reality, and how can it be validated?
2. **Methodological (or procedural):** Which principles and rules guide the process of inquiry?

This triad, originally systematised by Corbetta (2014), provides the conceptual grammar of scientific reasoning (see Figure 3.1). It underlies research paradigms across the natural and social sciences and defines the boundaries of what can be legitimately studied and known. Also Ding (2019) used the same group of questions, demonstrating how this framework can be operationalised for analysing quantitative research, showing that even apparently technical procedures rest on implicit answers to these three philosophical questions.

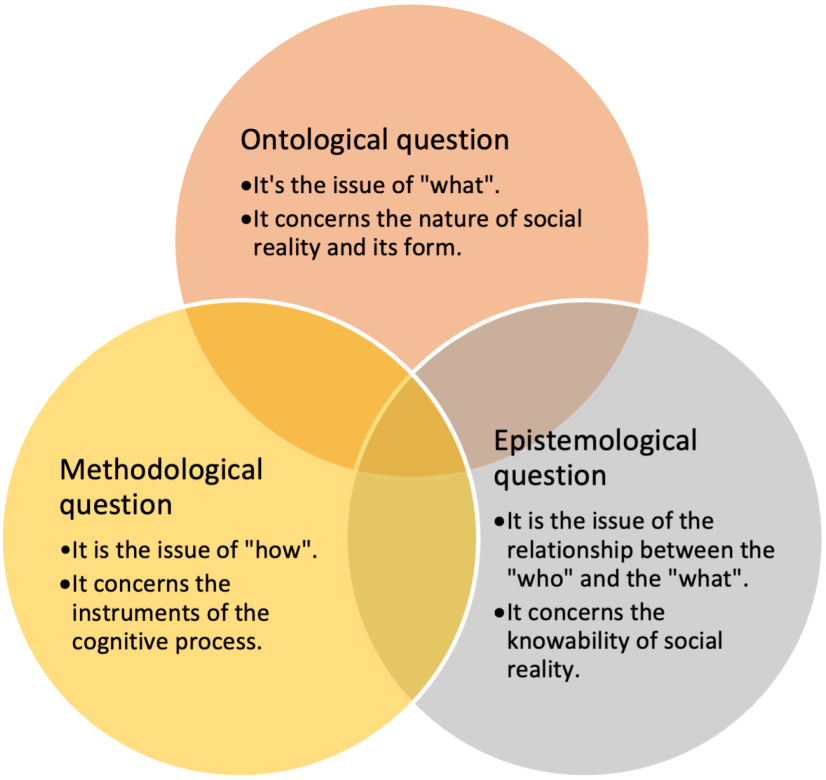

Figure 3.1 - Triad of foundational questions of scientific reasoning.

In this thesis, the triad functions as the conceptual backbone of the analytical tools. It offers a language for identifying and articulating the implicit foundations of diverse methods. However, while Corbetta and Ding emphasise ontology, epistemology, and methodology, the present work extends the framework to include a fourth, often neglected, dimension: axiology, the study of the values that orient scientific practice and knowledge production. This addition, grounded in Patterson and Williams (1998), responds to the increasing awareness that methods embody ethical and normative commitments such as transparency, reproducibility, or fairness. This dimension becomes important especially in the modern age, where methods are mediated by AI, and could privilege values of

efficiency at the expense of fairness. Axiology thus completes the triad, transforming it into a triadic-in-unity: ontology, epistemology, and axiology as interdependent dimensions of methodology.

3.3.3 Paradigms as worldviews in science education research

Within science education research, these philosophical dimensions are crystallised in the notion of *paradigm*. Treagust and Won (2023) define paradigms as coherent worldviews that determine what counts as valid knowledge, the goals of research, and the legitimate roles of researchers. Paradigms such as positivism, interpretivism, critical theory, and mixed methods each express distinct configurations of the ontological-epistemological-axiological triad.

However, as the field evolves under the influence of digitalisation and AI, these paradigms are increasingly fluid. The introduction of computational approaches has generated hybrid forms of inquiry that blend empirical, data-driven reasoning with interpretive and generative practices (Ganascia, 2010). In this emerging landscape, paradigms can no longer be treated as isolated camps but as dynamic constellations capable of dialogue. This pluralistic view provides the philosophical justification for developing tools that allow comparison across different methodological orientations, revealing where assumptions diverge and where common ground can be established.

The need for such pluralistic reflection echoes the long-standing philosophical debate on the *"sciences of nature"* and the *"sciences of spirit"* (Fantini, 2014). The former, rooted in positivism, pursues generalisable laws through observation and experiment; the latter, grounded in interpretivism, seeks understanding through contextual interpretation. These opposing traditions have historically shaped educational research, giving rise to the quantitative-qualitative divide.

Yet, contemporary methodological developments blur this dichotomy. AI-based methods occupy an *intermediate epistemic space*, combining empirical generalisation with interpretive modelling. They challenge the notion that explanation and understanding, deduction and induction, or measurement and interpretation belong to separate epistemologies. This shift invites a re-examination of the very paradigms that once defined scientific inquiry, demanding analytical frameworks flexible enough to accommodate hybridity.

3.3.4 Insights from practical experience

Alongside theoretical foundations, the development of the analytical tools was deeply informed by practical experience in data analysis. Across different case studies, it became possible to recognise a set of recurring analytical phases, such as data preparation, immersion, and reduction, modelling or exploration, coding, pattern development, and interpretation. Within these phases, several moments of *exchange* emerged: between human and machine actions, between externally injected knowledge and internally generated patterns, and between previous results and new lines of reasoning. These interactions revealed that knowledge does not simply accumulate, but evolves through a dynamic interplay of contributions and transformations.

Reflecting on these processes led to the formulation of the concept of *knowledge dynamics*, which is the continuous transformation of data into knowledge through iterative cycles of representation, interpretation, and validation. Each of these cycles entails multiple kinds of input: contextual understanding, theoretical frameworks, and algorithmic procedures. These distinct forms of

knowledge intersect and influence one another, often carrying implicit assumptions that shape both the process and its outcomes.

This recognition of the *shaping* of knowledge, through choices such as data reduction, moments of human-machine interaction, and the integration of external or contextual insights, highlighted the need for an analytical framework capable of distinguishing and articulating these intertwined dimensions.

3.4 Designing the comparative grids

The comparative grids were developed through a long process of ideation, testing, and refinement that unfolded over three major iterations, represented in Figure 3.2. Their conception stemmed from a practical need that emerged in the second case study, where we separately applied Thematic Analysis (TA) and Latent Semantic Analysis (LSA) to a large corpus of students' essays. I sought to create an instrument capable of reconstructing the methodological process, allowing the choices, implicit conditions, and distinctive ways in which each method treated data and generated themes or topics to become visible.

The first version of the grid was designed to enable a direct comparison between TA and LSA with respect to two main areas. The first concerned knowledge production effort, that is, the practical requirements that sustain the functioning of each method, such as software, analytical expertise, time, and the organisation of analytical phases. The second concerned the quality of the knowledge produced, articulated through dimensions that are meaningful across both qualitative and quantitative research traditions. These included:

- **Originality of topics**, meaning whether the method was capable of generating novel insights for the field or tended instead to reproduce known or trivial results;
- **Interpretability**, referring to the extent to which the patterns identified were meaningful and comprehensible to the researcher, a particularly relevant aspect given the challenges both methods encountered in different phases of analysis;
- **Generalisability**, which emerged as a key concern when the breadth of the sample limited the analyst's ability to review all essays in the TA, raising questions about the extent to which bottom-up qualitative coding could be extended to the whole corpus;
- **Applicability on the research**, in terms of how the outcomes shaped understanding and further analysis.

In addition, the grid considered the methodological constraints under which each process unfolded, highlighting the strengths and limitations intrinsic to both methods. This first version of the grid was presented at the 2023 conference of ESERA[6] in an oral contribution titled *"Epistemological implications of different methodological approaches in textual data analysis"*.

However, when revisiting the analysis for the subsequent book chapter among the selected ESERA contributions, I realised that this first attempt was disorganised and somewhat heterogeneous, with several overlapping categories. For example, *knowledge production effort* and *methodological constraints* often coincided, while the comparative structure tended to emphasise "pros and cons" rather than the deeper mechanisms through which methods operated. Moreover, the analytical process itself was not visible: the tool captured outcomes but not the evolution of the process that led to them.

[6] ESERA is the European Science Education Research Association (https://www.esera.org/)

In response, I proposed a second version, shifting the emphasis toward methodological phases, knowledge dynamics, and the roles and relationships among researchers, experts, and computational models. This version aimed to describe not only the products of analysis but also the epistemic movements through which data were transformed into knowledge. Yet, difficulties persisted. It proved impossible to describe, for example, the researcher's relationship with the LSA model or with data without simultaneously referring to the types of knowledge mobilised, by whom, and at what stage. Many dimensions remained intertwined, a natural reflection of the interconnectedness of methodological processes, but this interdependence made the tool complex to apply and difficult to interpret consistently.

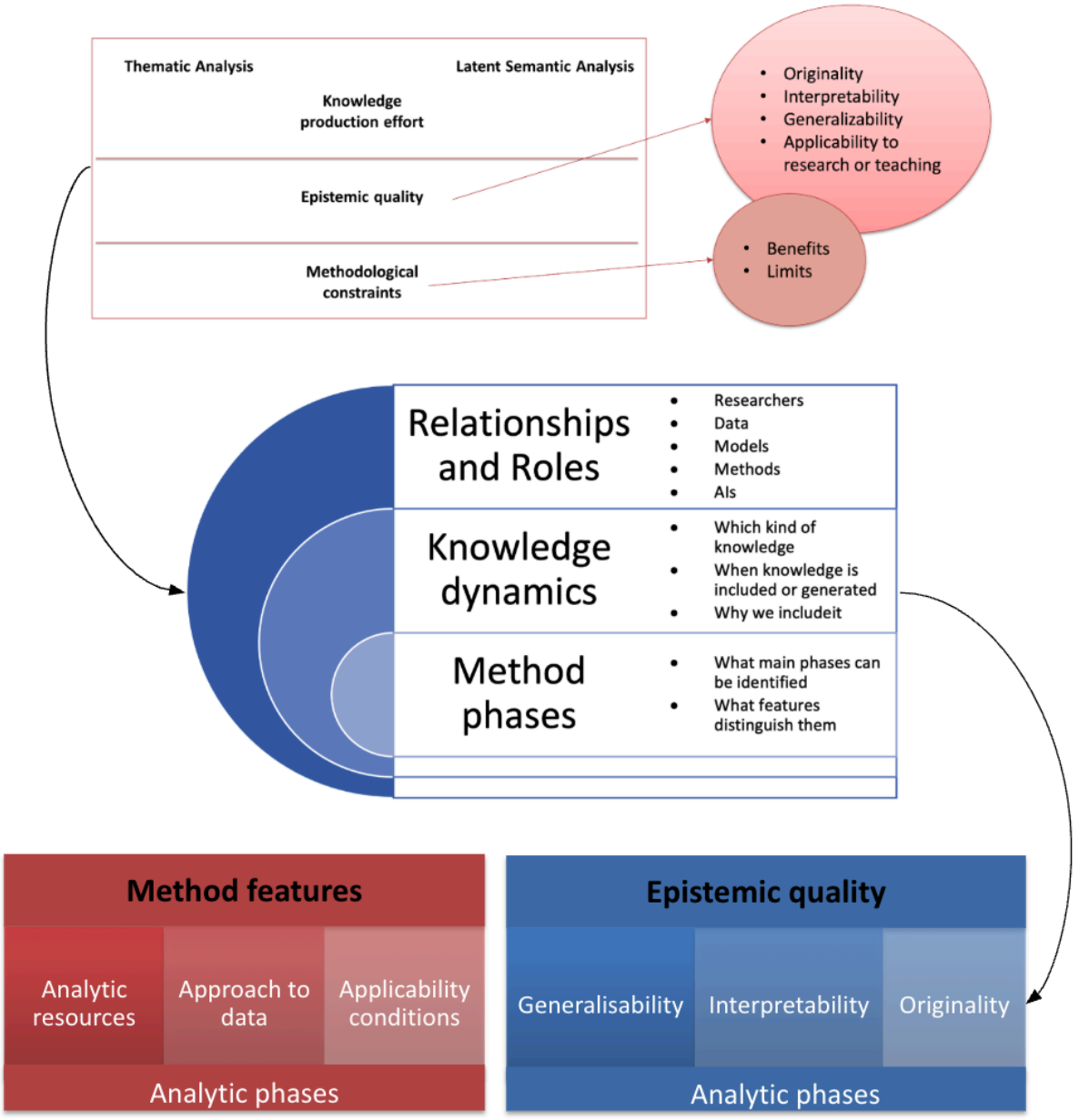

Figure 3.2 - Subsequent versions of the tools developed to analyse methods

At that stage, together with my supervisor, I decided to establish two guiding principles for further refinement, which ultimately led to the creation of two distinct tools:

- **Coherence and non-overlap principle:** each category should have a clearly delimited focus and stand independently, avoiding overlap with others.
- **Distinction of focus between method and methodology:** the analytical focus should be either on the *method* (e.g. how it operates, how knowledge is manipulated, and how the analytical process unfolds) or on the *methodology* (e.g. the assumptions, paradigms, and worldviews that justify and shape the method's logic).

Applying these principles required a process of conceptual cleaning. Several categories were abandoned or merged to ensure clarity and internal coherence.

The result of this refinement was the development of two complementary instruments:

A. The comparative grids, centred on *epistemic quality* and *method features*. These grids focus on the analytical process and the practical and epistemic characteristics of methods, thus on the way they operate on data and the kinds of knowledge they enable.
B. The heuristic tool, which addresses the dimension that had remained implicit until that point: the *why* behind methods. This second tool was conceived to explore the underlying logic of methods, that is the methodological foundations that give them meaning and legitimacy.

The next section presents this heuristic tool in detail, explaining its rationale, structure, and the visual and conceptual language chosen to represent it.

3.5 An heuristic tool to analyse methodological assumptions

3.5.1 Structure of the heuristic tool

The heuristic tool was conceived as a framework to make visible the philosophical logic that underlies research methods. As illustrated schematically in Figure 3.3, every methodology, understood as the logic that gives coherence and meaning to methods, is grounded in three foundational dimensions: ontology, epistemology, and axiology. These dimensions, though constitutive of any scientific practice, often remain implicit in everyday methodological discourse. The purpose of the tool is precisely to uncover these latent dimensions and to guide researchers in tracing how they manifest within the practical use of methods, including computational and AI-based ones.

The heuristic tool was designed to be also a meta-reflection tool, to be operative applied to the analysis of existing cases in literature. For this reason, it is organised into two components working over the three analytical lenses: *Procedur*e and *Guiding Questions*; and one general component, the *Markers* (see Table 3.1).

The first two components, *Procedure* and *Questions*, have a parallel function: they orient the researcher's attention toward identifying information and assumptions that reveal the ontological, epistemological, and axiological grounds of a method. For instance, by following the procedural steps or answering the corresponding guiding questions, the user is invited to examine what is considered knowable about the phenomenon under investigation when using a particular analytical approach, or which values are embedded in the way an algorithm, model, or procedure is designed and applied. In this sense, the tool functions as a reflective aid rather than an evaluative checklist: its aim is not to judge methods, but to make their internal logic explicit and discussable.

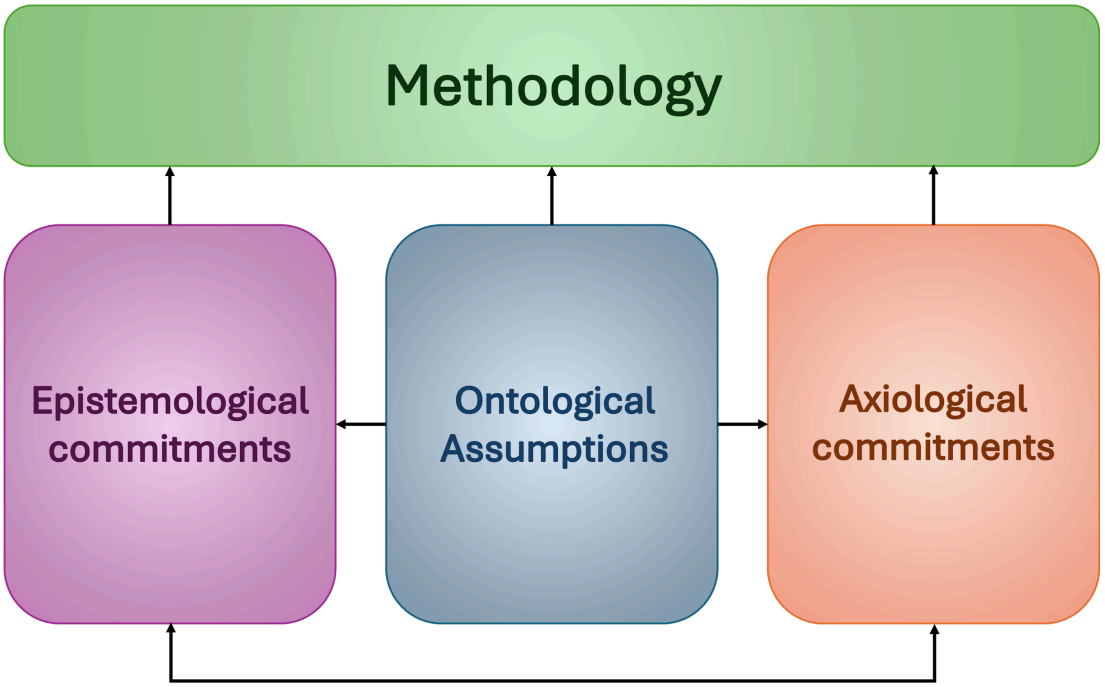

Figure 3.3 - A schematic organisation of the three main lenses of the heuristic tool, underlying methodology: ontological assumptions, and epistemological and axiological commitments.

The third component, the *Markers*, represents the distinctive contribution of this heuristic. Markers are those meaningful elements of a method that, through careful observation across different analytical experiences, proved capable of making a real difference between one methodological configuration and another. They act as *foci of attention* for the researcher, offering concrete entry points through which the implicit philosophical assumptions of a method can be recognised. The six markers identified are:

- Setting
- Researcher's role
- Data
- Previous contextual knowledge
- Theory behind the methods and embedded in the methodology
- Resources and method's practical role

These markers emerged from the prolonged reflection developed throughout the doctoral journey. Each time something implicitly significant about a method was observed, whether in its use of data, its treatment of context, or its distribution of roles between human and machine, it could be traced back to one of these six markers. For this reason, they were included in the final version of the heuristic tool, to serve as anchors for analysis and comparison.

Together, the three components constitute a structured path for uncovering the philosophical architecture of methods. They invite the researcher to move from the visible surface of methodological practice toward the deeper layers of its ontological, epistemological, and axiological foundations, thereby transforming methodological reflection into a deliberate act of inquiry.

Table 3.1 - Structure of the heuristic tool. It organises the heuristic around three foundational analytical lenses (ontology, epistemology, and axiology) each explored through a set of procedural prompts, guiding questions, and analytical markers. These components collectively support the identification of the implicit assumptions and values shaping how methods operate and generate knowledge.

Analytical lenses		**Procedure**	**Question**	**Markers**
Methodology	Ontology	Identify where the paper assumes or defines what is considered the object of analysis by the AI-based method, tool, or algorithm.	What is the object of the study that you analysed by using this AI-based method, tool, or algorithm? What is its "ontological status"?	➢ Setting ➢ Researcher's role ➢ Data ➢ Previous contextual knowledge ➢ Theory behind the methods and embedded in the methodology ➢ Resources ➢ Method's practice role
	Epistemology	Identify where the paper describes what is considered knowable about the object of analysis, or what is used as valid knowledge to study the object under study. Thus, determine what type of data the method relies on.	What do you consider to be knowable about the object of analysis when using this AI-based method, tool, or algorithm? What type of data does the method rely on?	
	Axiology	Identify where the paper discusses the values guiding the development of the method or those embedded in the knowledge it produces.	What values guided the application of the AI-based method, tool, or algorithm? What values are embedded in the knowledge produced by this method?	

3.5.2 A metaphor

The process of developing the heuristic tool was also accompanied by the search for a metaphor that could render its underlying logic more tangible and relatable across research traditions. Foundational reflections are never easy to introduce within established communities. Researchers often operate within familiar methodological environments, where shared practices and languages enable collaboration but may also create closed circles that view other approaches with suspicion. While such homogeneity fosters efficiency and coherence, it can also limit epistemic diversity. Learning from other perspectives, by contrast, can enrich the field and open it to new forms of inquiry. Yet, this dialogue is difficult when the assumptions behind methods remain implicit, overshadowed by technical routines that risk turning methodology into a matter of habit rather than reflection.

Seeking a metaphor that could invite a more conscious stance toward methodology, I initially explored several artistic and architectural inspirations. Works such as Zaha Hadid's *Vitra Fire Station*, Gehry's *Guggenheim Museum in Bilbao*, and Kandinsky's compositions of interlocking circles all offered elements of what I wished to convey: the tension between transparency and opacity, the harmony among distinct yet interconnected parts, and the invitation to look through and beyond surfaces (see Figure 3.4).

Figure 3.4 - Composition of inspirational images used during the heuristic tool's metaphor creation. Images reproduced for illustrative purposes only; no copyright ownership is claimed.

However, the image that ultimately crystallised my metaphor came from a personal recollection. During my doctoral period abroad in Oslo, I often visited the Oslo Opera House, a striking building designed by Snøhetta and situated at the edge of the fjord (see Figures 3.5 and 3.6). The Opera House rises from the water like a natural extension of the landscape, an architectural iceberg whose sharp lines and sloping surfaces suggest that much lies hidden beneath what is visible. Its vast rooftop, open to the public, allows people to walk freely across it without entering the theatre or even realising what lies inside (Figure 3.7 and Figure 3.8). Many, like myself, have experienced the building only from the outside, enjoying the view and the solid ground underfoot, without ever exploring its interior. Only later did I realise that this mirrored the way we often use research methods: we walk upon the surface of methodology, relying on its apparent solidity, without attending to the deeper structures that sustain it.

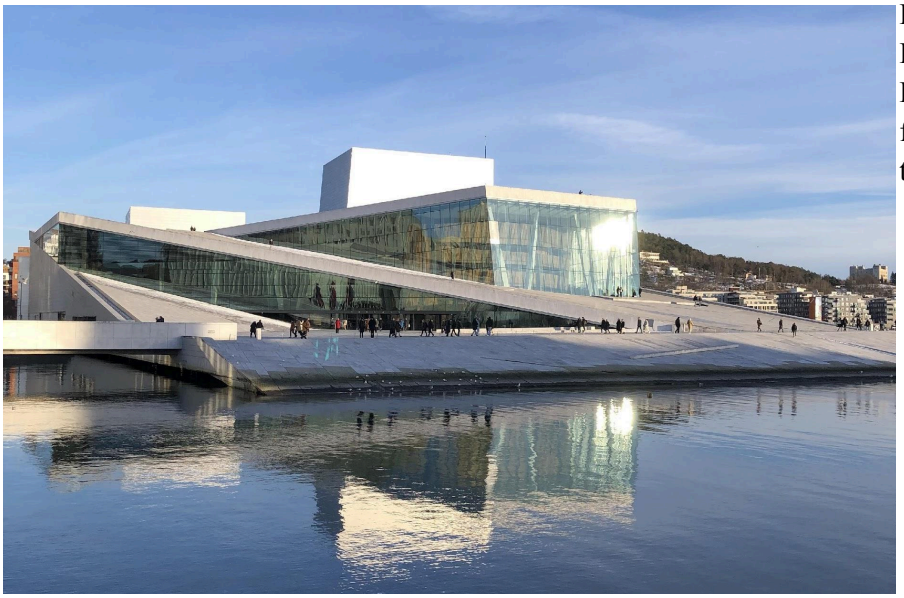

Figure 3.5 - Oslo Opera House (lateral view). Picture taken by me the first day I was in Oslo, the 31st of March, 2023.

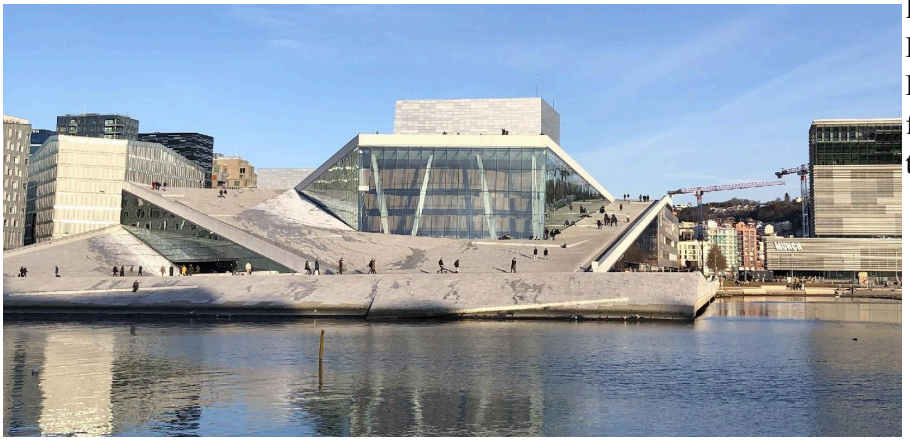

Figure 3.6 - Oslo Opera House (frontal view). Picture taken by me the first day I was in Oslo, the 31st of March, 2023.

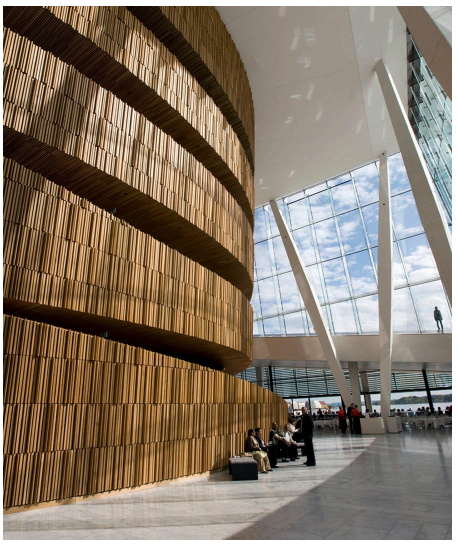

Figure 3.7 (left), Figure 3.8 (right) - Views of the Oslo Opera House from outside and inside. The large glass surfaces and the people walking on the roof symbolise researchers engaging with methods only at their visible level, often unaware of the deeper structure that sustains them. Inside, the materials of wood, metal, and stone evoke the ontological, epistemological, and axiological foundations of methodology. Images reproduced for illustrative purposes only; no copyright ownership is claimed.

The Opera House is built upon three primary materials, that are stone, wood, and metal, which together generate its distinctive coherence and balance (see Figure 3.7). In my metaphor, these correspond to the ontological, epistemological, and axiological dimensions of methodology. They are distinct yet inseparable, each contributing to the stability and aesthetic harmony of the whole. From the outside, their interplay is not immediately visible; only by entering the building or looking carefully through the glass walls does one perceive the order and resonance that bind them together.

In this sense, the Opera House captures the very essence of the heuristic tool, schematically represented in Figure 3.8. The visible roof represents the explicit level of methods and practices, the space where researchers walk, design, and operate. The hidden architectural structure beneath stands for the implicit foundations that give those practices meaning and coherence. The heuristic tool functions like an invitation to step inside: to look through the transparent walls of methodology, to explore what is usually unseen, and to recognise how ontology, epistemology, and axiology combine to sustain methodological integrity.

Just as the Opera House merges landscape, architecture, and experience into a unified form, the tool seeks to harmonise reflection and practice, helping researchers to balance technical rigour with philosophical awareness. It reminds us that methods do not stand on neutral ground: they rest upon deep structures, sometimes visible, sometimes concealed, that deserve to be visited, not merely walked upon.

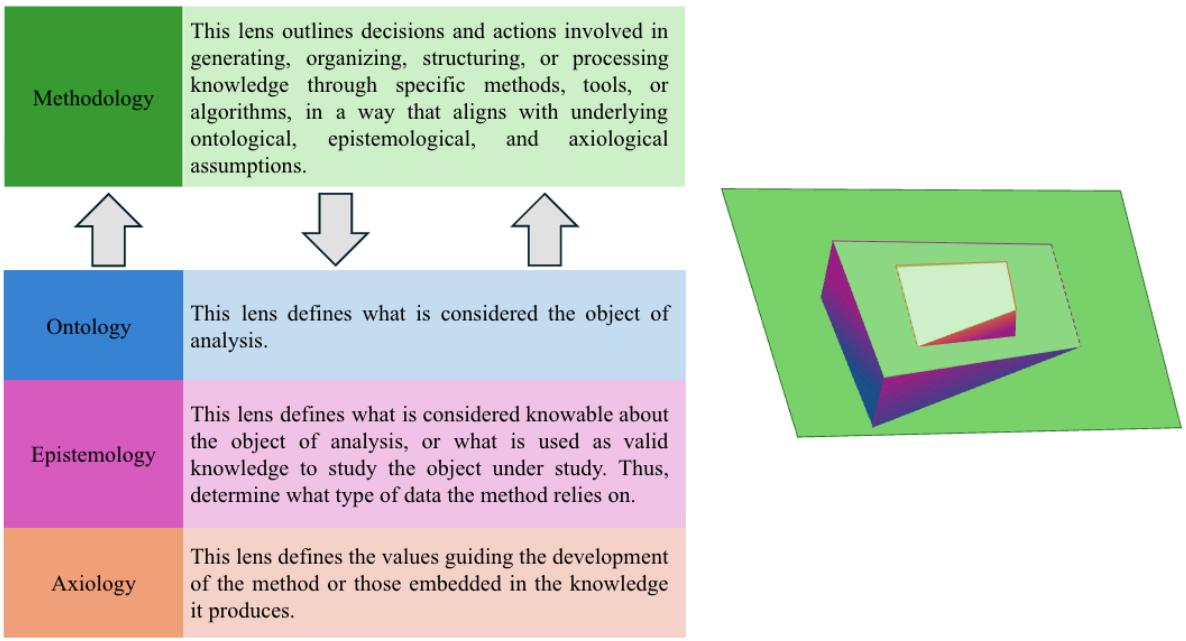

Figure 3.8 - Schematic and coloured confront between the heuristic tool and its metaphor, the Opera Oslo House.

3.6 Conclusions

This chapter marked a transition from the analytical exploration of individual methods to a meta-reflective inquiry into methodology itself. Building upon the comparative analyses developed in Chapter 2, it proposed a heuristic tool designed to make the philosophical and practical dimensions of

research methods more visible and comparable. The process of its construction, rooted in both theoretical reflection and empirical engagement, demonstrated that understanding how methods operate also requires understanding *why* they operate as they do.

The heuristic tool thus emerged as an evolution of the comparative grids, extending their analytical capacity from mapping methodological features to uncovering their ontological, epistemological, and axiological foundations. It offers a structured yet flexible support for researchers navigating across diverse methodological traditions, particularly when computational and qualitative approaches are combined. Rather than determining methodological hierarchies or prescriptive pathways, it encourages awareness of the assumptions, values, and forms of knowledge that shape research practice. By integrating procedures, guiding questions, and interpretive markers, the tool acts as an invitation to pause and reflect on the deeper architecture of methods, on the invisible structures that sustain their apparent solidity. In doing so, it contributes to a broader project of methodological transparency and dialogue within SER.

The next chapter takes up this invitation by examining how AI-based methods are transforming the very logic of scientific discovery. Through the application of the heuristic tool, the reflection initiated here finds a broader resonance, illuminating how the ongoing evolution of methods in science can inform, challenge, and ultimately enrich methodological thinking in SER.

References

Corbetta, P. (2014). Metodologia e tecniche della ricerca sociale (2nd ed.). Il Mulino.

Ding, L. (2019). Theoretical perspectives of quantitative physics education research. Physical Review Physics Education Research, 15(2), 020101. doi: 10.1103/PhysRevPhysEducRes.15.020101

Fantini, P. (2014). *Verso una teoria locale dell'appropriazione nell'insegnamento/apprendimento della fisica*. PhD dissertation.

Ganascia, J. G. (2010). Epistemology of AI Revisited in the Light of the Philosophy of Information. *Knowledge, Technology & Policy,* 23, 57–73. https://doi.org/10.1007/s12130-010-9101-0

Henderson, C. (2017). Editorial: Call for Papers for Focused Collection of Physical Review Physics Education Research Quantitative Methods in PER: A Critical Examination, *Physical Review Physics Education Research,* 13, 010002. doi: 10.1103/PhysRevPhysEducRes.13.010002.

Henderson, C. (2021). Editorial: Call for Papers for Focused Collection of Physical Review Physics Education Research Qualitative Methods in PER: A Critical Examination, *Physical Review Physics Education Research,* 17, 020001. doi: 10.1103/PhysRevPhysEducRes.17.020001.

Patterson, M. E., & Williams, D. R. (1998). Paradigms and problems: The practice of social science in natural resource management. *Society & Natural Resources*, *11*(3), 279–295. https://doi.org/10.1080/08941929809381080

Pietsch, W. (2015). Aspects of theory-ladenness in data-intensive science. *Philosophy of Science* 82.5, 905-916. https://doi.org/10.1086/683328

Schwandt, T. A. (2007). Dictionary of qualitative inquiry (3rd ed.). Sage.

Treagust D. F., & Won, M. (2023). Theory and Methods of Science Education Research. In Boone, W. (Eds.), Handbook of Research on Science Education: Volume III (pp. 3-27). Routledge.

Zanellati, A., Mitri, D. D., Gabbrielli, M., & Levrini, O. (2024). Hybrid Models for Knowledge Tracing: A Systematic Literature Review. *IEEE Transactions on Learning Technologies*, 17, 1021-1036. https://doi.org/10.1109/tlt.2023.3348690

Chapter 4 – AI in science

Mapping a methodological shift

4.1 Introduction

Following the development of a heuristic tool for methodological reflection in Chapter 3, this chapter extends the inquiry by focusing on AI-based methods as they emerge and evolve within contemporary scientific research. The aim is to deepen our understanding of the philosophical and methodological assumptions embedded in these approaches. In doing so, the chapter examines the novelties they introduce, recognising that AI is not merely accelerating the scientific process but transforming how science itself is conducted.

To this end, after introducing scientific methods in science education, how they are presented in school, and the theoretical framework of reference (Section 4.4), two milestone publications illustrating how methodologies in science are changing were analysed. The *Nature* article *Scientific Discovery in the Age of Artificial Intelligence* (Wang, et al., 2023) was used to map out current AI-based methods across scientific domains (Section 4.5), while *Highly Accurate Protein Structure Prediction with AlphaFold* (Jumper, et al., 2021) was examined through the heuristic tool introduced in the previous chapter (Section 4.6). As an AI tool that led to the 2024 Nobel Prize in Chemistry, AlphaFold serves as a concrete case illustrating how AI integrates traditional approaches into new hybrid epistemic strategies, transforming how knowledge is developed and treated. Together, these studies offer a window into how AI-driven methods are conceptualised, implemented, and justified within the scientific community.

By mapping the novelties brought by AI within the broader evolution of scientific methods and practice, this chapter contributes to science education, offering educators and researchers a critical lens to interpret the ongoing transformation of scientific methods and their epistemic foundations. Also, it contributes to the general objective of this thesis, providing insights from the scientific research field, now increasingly shaped by automation, large-scale data processing, and algorithmic reasoning. Similarly science education research (SER) hears the echo of analogous methodological changes in the current era of AI.

4.2 Situating a study on Science within the thesis framework

In recent years, AI has become a major driver of both scientific and societal transformation, reshaping not only how science is conducted but also how it is taught, learned, and researched (Erduran, 2023). AI systems now intervene in multiple stages of scientific discovery (from hypothesis generation and experimental design to data interpretation) enabling forms of inquiry that were previously unimaginable with traditional methods (Wang et al., 2023). As a result, science itself is undergoing a profound methodological and epistemological shift.

Parallel developments can be observed within SER, where AI serves as a tool for teaching and learning, a topic of instruction, and a research instrument (Cooper, 2023; Jia et al., 2024; Billingsley et al., 2025; Wulff et al., 2025). However, beyond these functional roles, scholars such as Erduran (2023) and Erduran and Levrini (2024) have argued that AI is also transforming the very Nature of

Science (NOS), that is, the fundamental characteristics and values that define scientific inquiry as a distinctive way of understanding the world. The call to revisit NOS highlighted the need to reassess what we teach about science in light of these transformations, noting that as AI reshapes scientific reasoning and practice, education must evolve to reflect these changes.

Building on this call, the present chapter broadens the focus of the thesis. While the previous chapters developed and tested a heuristic tool for methodological reflection within SER, the same kind of methodological reasoning and transformation observed there is now clearly visible in scientific research itself. The decision to apply the heuristic tool beyond education to the domain of science, arose from this convergence of changes: both science and science education are being reconfigured by AI-driven, data-intensive methods. The tool was thus ready at a moment when such reflection became most needed, allowing us to contribute to the ongoing NOS debate from an educational-philosophical standpoint. Accordingly, this chapter examines how AI-based methodologies in science are reshaping the epistemological, ontological, and axiological assumptions that underpin scientific inquiry. These methods are not neutral: they embody specific worldviews about what counts as valid knowledge, what entities are knowable, and which values govern scientific reasoning (Ding, 2019). In this sense, understanding AI's impact on scientific methodology is essential for anticipating how science education, and particularly the teaching of NOS, should adapt.

The following sections articulate the theoretical background (Section 4.3), specify the aims and research questions (Section 4.4), and present the analytical work conducted on AI-based scientific methods, including the AlphaFold case study (Sections 4.5 - 4.7). The chapter concludes by linking the findings back to the broader methodological argument of the thesis and to current discussions on NOS and science education (Sections 4.8 - 4.9). This study has also been developed in parallel as a manuscript submitted to *Science & Education* journal (Caramaschi, Levrini & Erduran, submitted), where its main findings are presented in article format. The present version expands and contextualises that work within the theoretical and methodological framework of this thesis.

4.3 Aims, research questions, and argumentation

The overarching aim of this study is to contribute to the ongoing discussion on the NOS within the field of science education, in light of the transformations introduced by AI. More specifically, this chapter seeks to investigate how AI-based approaches are reshaping the methodological foundations of scientific research, and what these changes imply for the way NOS is conceptualised, taught, and learned.

To guide this inquiry, two research questions have been formulated:

- RQ1. What examples of epistemological, ontological, and axiological novelties emerge from the use of AI-based methods in scientific research?
- RQ2. What implications do these methodological shifts have for the teaching and learning of NOS, particularly regarding scientific methods and methodological rules?

The argumentative approach adopted to answer these questions parallels the study of Erduran and Levrini (2024), published in the International Journal of Science Education. Basically, we adopt a theoretical orientation, but drawing upon examples from empirical scientific research to illuminate the methodological shifts introduced by AI. Also, the inquiry is grounded in the same theoretical framework on NOS employed in previous chapters, focusing on the dimension of scientific methods

as a key component of what Erduran and Dagher (2014) describe as the *cognitive-epistemic system* of science. By analysing how AI influences this dimension, the chapter aims to extend the methodological reflection initiated within SER to the broader context of scientific practice, thereby contributing to current efforts to reconceptualise NOS in the era of AI.

4.4 Scientific methods in science education

4.4.1 The Nature of Science Framework

The analysis of scientific methods undertaken in this chapter is grounded in the broader field of research commonly referred to as the Nature of Science (NOS) (Lederman et al., 2002; Erduran & Dagher, 2014; Irzik & Nola, 2014). While NOS has been extensively discussed in the philosophy of science, establishing a unified and universally accepted definition remains an open challenge (Irzik & Nola, 2011). Similarly, within the science education community, multiple interpretations coexist, and there is still limited consensus on which aspects of science should be prioritised in teaching and learning (Osborne et al., 2003; Erduran & Dagher, 2014).

Over time, several conceptual frameworks have been developed to characterise NOS. The consensus view (Lederman, et al., 2002) has been particularly influential for its pedagogical clarity and its ability to translate philosophical ideas into teachable components. However, other models such as the Family Resemblance Approach (FRA) to NOS (Irzik & Nola, 2011; Erduran & Dagher, 2014) and Whole Science (Allchin, 2017) have expanded the field by offering more integrative and dynamic perspectives on science.

Among these, the FRA framework provides the theoretical foundation adopted in this study (Figure 4.1). Developed by Irzik and Nola (2011, 2014) and subsequently elaborated for science education by Erduran and Dagher (2014), FRA offers a holistic and flexible account of science. Drawing inspiration from Wittgenstein's notion of "family resemblances", it avoids prescribing a rigid set of necessary and sufficient features that all sciences must share. Instead, it seeks to capture the diversity and overlap among different scientific disciplines (e.g., chemistry, biology, physics) by mapping commonalities across several interrelated dimensions, epistemic, cognitive, institutional, and social.

This framework is particularly suitable for the present study for two main reasons. First, it allows for the integration of multiple dimensions of science within a single, coherent model that remains open and adaptable to change, an essential quality given the rapid transformations currently occurring in the age of AI. Second, it explicitly distinguishes between scientific practices and scientific methods, a distinction that aligns with the focus of this chapter on methodological transformation.

In the inner system of the FRA "wheel" (Erduran & Dagher, 2014), four key categories are identified: *aims and values*, *methods and methodological rules*, *scientific practices*, and *scientific knowledge*. Together, these constitute the cognitive-epistemic system of science. The present analysis concentrates on the category of *methods and methodological rules*, given its direct relevance to our investigation of how AI-based approaches are reshaping the logic of inquiry and the criteria for knowledge validation. Importantly, the FRA treats each category as dynamic and historically situated, acknowledging that scientific methods evolve alongside technological, social, and epistemic developments. This flexibility is crucial for understanding the fluid and rapidly changing nature of AI-driven research methods.

Within the FRA framework, scientific practices and scientific methods are conceptually distinct yet interconnected. *Practices* encompass the range of activities scientists engage in, such as modelling, classifying, observing, or experimenting, each influenced by existing theories, values, and the socio-political context of scientific work. For instance:

- *Observation* becomes a scientific practice when embedded within theoretical frameworks and linked to other epistemic activities such as modelling or inference.
- *Classification* acquires scientific meaning when it serves explanatory or predictive purposes.
- *Experimentation* is not a fixed procedure but evolves in relation to the tools, data, and aims guiding inquiry.

By contrast, *methods and methodological rules* refer more specifically to the strategies, criteria, and norms that govern how knowledge is produced, validated, and accepted within the scientific community. These include methodological standards ensuring coherence, reliability, and conformity to disciplinary expectations. This distinction is essential for the present study: it enables a more precise analysis of how AI is reconfiguring not only what scientists do (their practices) but how they justify and legitimate their work (their methods). Consequently, the FRA framework provides the conceptual scaffolding needed to investigate the evolving nature of scientific methodology, and to interpret what these transformations mean for how NOS should be understood and taught in contemporary science education.

Figure 4.1 - FRA wheel: science as a cognitive-epistemic and social-institutional system (Erduran & Dagher, 2014, p.28)

4.4.2 Scientific methods in science education

The teaching and learning of scientific methods has long been a central objective in science education. However, the ways in which "the scientific method" is typically presented in classrooms and textbooks have been widely criticised for oversimplifying and misrepresenting how science actually works. Educational materials often portray scientific inquiry as a linear, stepwise process that is assumed to apply uniformly across all scientific domains (Woodcock, 2013). Typically, this process is represented as made of observation, hypothesis formation, experimentation, data analysis, and conclusion. However this depiction, though pedagogically convenient, fails to reflect the diversity of actual scientific practice. As scholars have noted, the methodological procedures of different disciplines vary considerably; for example, fields such as geology or astronomy rely heavily on historical and interpretive approaches rather than experimental manipulation (Frodeman, 1995).

In reality, scientists engage in disciplined inquiry that involves a wide range of observational, investigative, and analytical strategies. These are guided by specific methodological rules that ensure the reliability and validity of scientific knowledge (Yeh, et al., 2019). As articulated in the FRA to NOS (Erduran & Dagher, 2014), scientific methods and methodological rules together form a core epistemic component of science, shaping how evidence is generated and how knowledge claims are justified in many diverse ways.

Despite increasing recognition of this complexity, science education has traditionally relied on a single, canonical model of the scientific method. The persistence of this model is partly due to its clarity and structure, which make it appealing for instructional purposes. Yet, as Ioannidou and Erduran (2021) and others have argued, this representation neglects the pluralism that characterises real scientific practice and can inadvertently convey a misleading image of science as a uniform and mechanical process.

To counter this reductionist view, science education research has increasingly advocated for a pluralistic and authentic portrayal of scientific methods in school curricula (NGSS, 2013; Wivagg & Allchin, 2002). This line of research encourages educators to move beyond "the" scientific method and instead highlight the diversity of approaches that scientists employ in different contexts. Nevertheless, the linear model remains deeply embedded in educational practice, particularly in introductory textbooks and laboratory instruction, where it continues to serve as a simplified heuristic for inquiry.

In response to these limitations, various conceptual frameworks have been proposed to represent methodological diversity more accurately. One influential example is Brandon's Matrix (Brandon, 1994), introduced to science education by Erduran and Dagher (2014). The matrix positions methods along two dimensions: whether they involve manipulation of variables and whether they are driven by hypothesis testing. This creates a continuum or "space of experimentality," demonstrating that not all scientific methods are experimental, and not all are based on testing hypotheses (see Figure 4.2 and Figure 4.3).

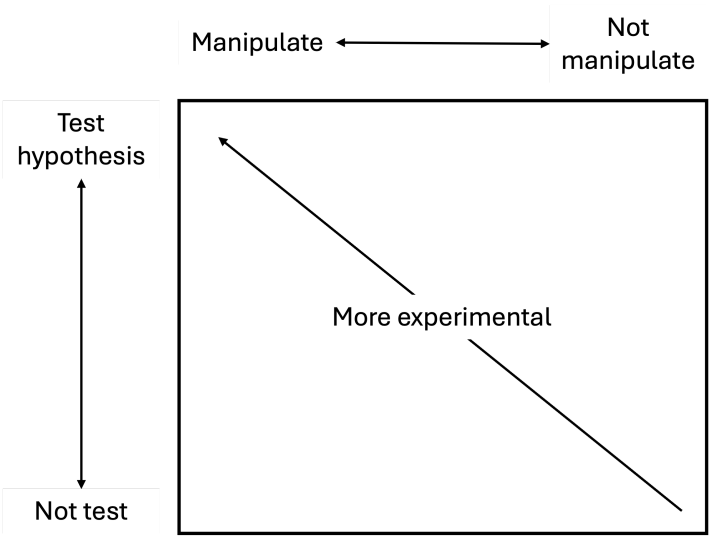

Figure 4.2 - Brandon's representation of "space of experimentality" between two continua (image reproduced from Brandon 1994, p.66)

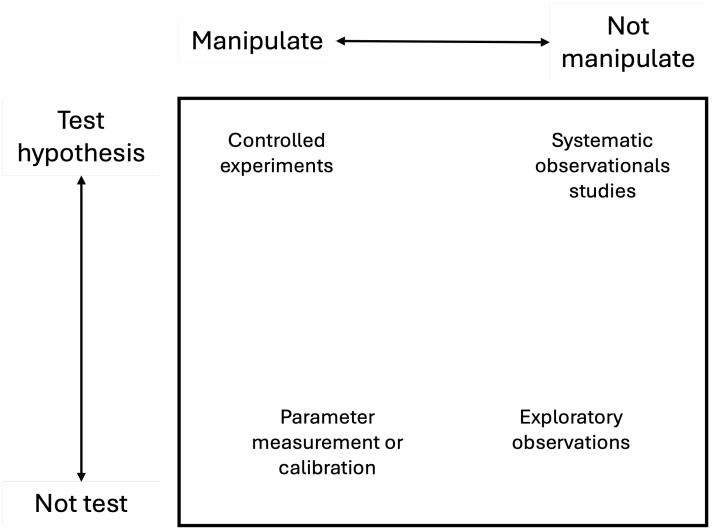

Figure 4.3 - Examples of methods in science positioned in the continuum of Brandon's matrix.

Building on Brandon's representation, a growing body of science education research (e.g., Chen et al., 2024; Cullinane et al., 2019; Erduran & Dagher, 2014; Ioannidou & Erduran, 2021; Liu et al., 2024; Upahi et al., 2025) emphasises the importance of teaching students about a range of valid scientific methods. Examples include:

- Controlled experiments, which involve both variable manipulation and hypothesis testing in tightly regulated conditions;
- Parameter measurement and calibration, where variables are manipulated to collect data or optimise conditions without testing formal hypotheses;
  Systematic observational studies, which are hypothesis-driven but non-manipulative, as in behavioural or ecological research; and
- Exploratory observations, which involve neither manipulation nor hypothesis testing, and are often used in the early stages of research to generate new questions or patterns.

These examples illustrate that scientific inquiry encompasses a continuum of methodological strategies, ranging from experimental to observational, and from hypothesis-driven to exploratory (Irzik & Nola, 2014). Recognising this diversity provides a more accurate and dynamic picture of how science operates in practice and helps challenge the persistent myth of a single, uniform "scientific method."

Equally important, scientific methods are always embedded within methodological rules which are the conventions, norms, and standards that govern appropriate use of particular methods in specific disciplines (Erduran & Dagher, 2014; Irzik & Nola, 2014). These rules not only ensure the rigour and reliability of knowledge production but also reflect the values and ethical principles underpinning scientific inquiry. Making these dimensions explicit in science education enables students to understand that science is both methodologically pluralistic and guided by principled reasoning, fostering a deeper and more authentic understanding of what it means to do science.

4.5 Mapping AI-based scientific methods in modern Science

Since there is no official or comprehensive classification of AI methods currently employed in scientific research, we decided to develop a mapping to identify and describe the main approaches through which AI contributes to scientific inquiry. To do so, we used the influential *Nature* article *"Scientific Discovery in the Age of Artificial Intelligence"* (Wang et al., 2023) as a central reference. The article offers a cross-disciplinary synthesis of how different AI techniques (e.g. self-supervised learning, geometric deep learning, generative modelling, reinforcement learning, and NLP) are being integrated into research across multiple scientific domains, including biology, chemistry, physics, and materials science.

Each section of the article was carefully examined, and every explicit reference to an AI method, tool, or technique was extracted. These were then coded and grouped according to their scientific function within the research process. In some cases, subcategories were created when multiple techniques contributed to the same stage of inquiry. Illustrative examples (for instance, geometric deep learning applied to protein folding or generative models for molecule design) were recorded to substantiate the categorisation. The resulting synthesis is summarised in Table 4.1, while the complete mapping is provided in Appendix A.

Table 4.1 - Summary of AI-based methods described in *"Scientific discovery in the age of artificial intelligence"* (Nature, 2023). This table shows key AI-based methods, including a description and examples in science, as well as scientific functions.

Scientific function	AI technique / category	Specific methods	Description	Examples in science
Data collection & curation	Deep Learning / Supervised Machine Learning / Graph-based ML / Bayesian Statistics / Generative AI	Multimodal anomaly detection and labeling with weak supervision	Uses pretrained models and active sampling to label large scientific datasets efficiently from noisy, heterogeneous data	Labeling cell types in single-cell RNA-seq; labeling satellite or microscopy images

Data representation	Graph Neural Networks (GNNs) / Deep Learning / NLP	Learning on structured scientific domains	Learns over molecules, materials, or equations represented as graphs; captures structural and relational information	Molecular property prediction; material design
Hypothesis generation	Symbolic regression + Reinforcement Learning	Sequential generation of symbolic laws	Uses neural policies to construct mathematical expressions that fit data, optimizing via reward signals	Discovery of physical laws from motion data; mathematical conjecture generation
Hypothesis generation	Deep Generative Models (VAEs)	Latent space optimization of candidates	Encodes hypotheses (e.g. molecules) into differentiable space; optimizes properties via gradient descent; decodes to concrete candidates	Protein or drug design from learned molecular representations
Experimentation	Active Learning + Bayesian optimization	Experiment selection under uncertainty	Selects the most informative or promising experiments based on model uncertainty and expected improvement	Ligand screening; protein variant testing; accelerator tuning
Simulation	Neural Partial Differential Equations (PDE) Solvers / Physics informed Neural Networks	Learned solution of physical equations	Trains networks to approximate PDE solutions using physics constraints, accelerating simulation	Climate models; fluid dynamics; seismic wave propagation

The analysis revealed a wide variety of AI techniques currently used across the sciences, including deep-learning architectures, graph neural networks (GNNs), Bayesian and probabilistic models, reinforcement learning systems, generative AI models, and NLP pipelines. These approaches support a range of research activities, such as detecting rare events, generating new molecular or material candidates, encoding geometric structures, and optimising experimental conditions.

To capture their functional diversity, the identified methods were organised into four major clusters, corresponding to their contribution to different stages of the research process:

1. **Data collection and curation**: automated data acquisition, weak labelling, anomaly detection.
2. **Data representation and transformation**: graph-based representations, vector encoders, geometric deep learning.
3. **Hypothesis generation and model inference**: pattern identification, probabilistic and self-supervised learning for model construction.
4. **Experimentation and simulation**: reinforcement-learning systems and surrogate models guiding or automating experimentation.

Across all categories, data emerged as the epistemic core of AI-based scientific methods. The *Nature* article highlights that AI's effectiveness relies on the digitisation of research and the abundance of structured and unstructured datasets. Through massive data resources, AI systems detect patterns, model relationships, and generate predictions that extend beyond the reach of traditional experimental approaches.

At the same time, the analysis highlighted the fragility of scientific data. Datasets are often incomplete, noisy, or heterogeneous, demanding extensive cleaning, integration, and validation before they can be used effectively. This fragility contrasts with the promise of automation and speed commonly associated with AI: in practice, the initial stages of data preparation involve a high degree of human attention, judgment, and decision-making regarding how data are selected, processed, and represented. The genuine automation of analysis typically begins only once the data are ready, meaning that human expertise remains essential for ensuring their reliability and epistemic integrity. Many AI methods directly address these challenges by generating synthetic data or pseudo-labels thereby augmenting existing datasets and enhancing their usability. In this sense, AI participates not only in analysing data but also in producing and refining it, an important epistemological shift in the logic of scientific discovery.

A central finding of this analysis concerns the evolving role of methodological rules in AI-based science. The *Nature* article reports multiple cases in which AI systems exhibit autonomous methodological behaviour, for example, adjusting experimental parameters, refining models in real time, or selecting the most relevant features for analysis. These systems act according to procedural rules that define how data are processed, validated, and interpreted. Traditionally, such rules served to guide human researchers, ensuring coherence, reproducibility, and adherence to disciplinary standards. In AI-driven research, however, they also govern the functioning of the AI systems themselves. Methodological rules thus provide a shared framework that structures how both humans and AI agents interpret data and make decisions throughout the scientific process. This dual governance marks a new form of distributed autonomy in scientific inquiry. Decision-making is increasingly shared between human and machine agents, coordinated through the same normative principles. Methodological rules become mediating entities, a common epistemic layer that sustains alignment, accountability, and consistency across human-AI collaborations. This finding represents a major novelty of the mapping, revealing how the notion of "method" is being redefined as methodological reasoning itself becomes encoded and operationalised within intelligent systems.

4.6 The AlphaFold case study

4.6.1 Overall methodology and operationalisation of the heuristic tool

The preceding discussion outlined the broader context of how AI is reshaping scientific methods, particularly regarding the management, interpretation, and generation of data. This section explains how the heuristic tool developed in Chapter 3 was operationalised to deepen AI-based methods foundations to identify the transformations introduced by AI in the logic of scientific inquiry.

I conducted a qualitative analysis on a landmark case study: the *AlphaFold* method, as described in *Highly Accurate Protein Structure Prediction with AlphaFold* (*Nature*, 2021). This article was selected for its exceptional scientific impact, its methodological transparency in describing a breakthrough AI-based approach in structural biology, and its recognition through the 2024 Nobel Prize in Chemistry. This article offers a detailed and technically rich account of a computational system that redefines traditional experimental practices, making it particularly suitable for analysing AlphaFold's methodological assumptions. This approach aligns with studies that have adapted theoretical frameworks to analyse empirical material, such as the use of Toulmin's Argument Pattern for investigating students' reasoning in science education (Erduran et al., 2004).

The analysis, based on the principles on which the heuristic tool was developed, builds on the distinction between method and methodology outlined in Chapter 3. While the former refers to the concrete investigative procedures used in research, the latter concerns the underlying principles and rules that govern how such procedures are designed, justified, and integrated within a broader system of knowledge production. To apply the tool, I conducted a close reading of the selected *Nature* article, focusing specifically on passages that reflected one or more of the three analytical lenses central to the heuristic framework (ontological, epistemological, axiological assumptions).

During this analytical reading, I highlighted all segments of the text that explicitly or implicitly addressed the ontological, epistemological, or axiological dimensions of inquiry. To ensure consistency and rigour, the process was informed by the markers and guiding questions defined in the heuristic tool (Chapter 3). While the markers served as *signposts* for recognising when methodological reasoning was made visible. For example, text was highlighted when referred to decision-making procedures, evaluative criteria, or the relationship between data and theory. Instead, the guiding questions functioned as *interpretive prompts* to focus the analysis. For instance, these questions asked:

- *What is the object of study analysed through this AI-based method, tool, or algorithm?*
- *What is its "ontological status", that is what kind of entity is considered knowable or real?*
- *What do we claim to know about this object when using this AI-based approach?*
- *What types of data does the method rely on, and how are they treated as evidence?*
- *Which values or guiding norms shape the application of the method?*
- *Which values are embedded in the knowledge it produces?*

Engaging with these questions while reading helped maintain attention on the core object of analysis within each passage, continually asking, *is this section referring to the object under study, to what is known about the object, or to the values and norms that orient the methodological process?* This iterative questioning ensured that the extraction of information from the text was analytically coherent

and aligned with the three dimensions of the heuristic framework, while also keeping the focus on how AI-based methods embody specific assumptions, epistemic commitments, and value orientations.

Given the **t**echnical complexity of the AlphaFold paper, ChatGPT was employed as a support tool. It both served to understand difficult technicalities about the method Also, the AI system was used to facilitate the preliminary extraction of potentially significant passages; however, all outputs were carefully reviewed, contextualised, and critically interpreted by the researcher to ensure accuracy and analytical rigour. The integration of AI in this phase therefore served as a practical aid, not a substitute for human interpretation, maintaining full scholarly oversight throughout the process.

4.6.2 AlphaFold: the method

AlphaFold represents a landmark achievement in contemporary science, not only because of its groundbreaking results but also because of its profound impact on how scientific research is conducted. The *Nature* article *"Highly Accurate Protein Structure Prediction with AlphaFold"* (2021) was selected as a case study because it exemplifies a widely recognised, AI-driven scientific method, an innovation acknowledged with the 2024 Nobel Prize in Chemistry. This shared recognition by the scientific community underscores the method's significance and its validity as a transformative milestone in computational biology.

The method addresses one of the longest-standing challenges in modern science: predicting the three-dimensional structure of proteins solely from their amino acid sequences. Proteins are fundamental to all biological processes, and understanding their structure is essential for elucidating their function. Despite decades of experimental efforts that have resolved the structures of approximately 100,000 unique proteins, this represents only a small fraction of the billions of known protein sequences. Experimental determination of structures remains time-consuming (often requiring months or years for a single protein) creating a major bottleneck in biological research (Jumper et al., 2021).

Consequently, accurate computational approaches have become essential to bridging this gap and enabling large-scale structural bioinformatics. The so-called "protein folding problem", the task of predicting a protein's structure from its sequence, has remained a central unsolved question for more than five decades. Previous computational methods had achieved partial success, but typically fell short of atomic-level precision, particularly when no homologous structures were available to guide predictions (Dill et al., 2008; Senior et al., 2019; Jumper et al., 2021).

The developers of AlphaFold combined existing bioinformatics strategies with a redesigned neural network architecture capable of achieving near-atomic accuracy in predicting protein structures. Remarkably, the system also succeeded for proteins lacking homologous templates, a breakthrough widely recognised during the 14th Critical Assessment of Protein Structure Prediction (CASP14), where AlphaFold's predictions rivalled experimental data and dramatically outperformed competing models.

4.6.3 Results of the AlphaFold analysis

Ontological assumptions

The ontological analysis draws upon selected excerpts from the *Nature* article (labelled from O1 to O6 in Table 4.2) accompanied by interpretive commentary. The analysis reveals a dynamic negotiation between two co-existing conceptualisations of the object of study, the protein structure, here referred to as computational ontology and biological ontology.

At the core of AlphaFold's approach lies a computational reinterpretation of the protein. Rather than being treated in its full biological complexity, the protein is represented as a geometric system composed of atom-level coordinates and backbone transformations (as seen in excerpts O1 and O3). In this view, the protein becomes a computable and manipulable entity, reducible to numerical relationships, something a machine can process efficiently.

This abstraction deepens in O2, where protein folding is reformulated as a graph inference problem in three-dimensional space. Amino-acid residues are treated as nodes, and their spatial proximities as edges, establishing a purely structural, topological view of the molecule. The biological materiality of the protein is thus set aside in favour of a computationally tractable abstraction. These modelling decisions are not merely technical choices; they reveal a profound ontological reframing of what the protein *is*, from a biochemical substance to a relational, data-driven construct.

Although biology has long relied on computational representations of molecular structures, in AlphaFold this abstraction is not merely convenient but constitutive of the scientific process itself. The method depends on rendering the object of study as a numerical, manipulable, and relational entity, whose behaviour can be predicted through data patterns. Without this ontological transformation, the protein could not be processed by the neural network, nor could knowledge be generated through this approach.

Yet this computational ontology is not absolute. As indicated in O4, AlphaFold temporarily permits deviations from strict biochemical constraints during prediction, allowing the model to explore intermediate configurations that may not correspond to physically realistic states. This flexibility constitutes a methodological asset, enabling the system to optimise its internal parameters. Nevertheless, this suspension of biological realism is temporary. As described in O5, the final predicted structures undergo post-prediction relaxation using the Amber force field—a simulation tool that re-imposes physical plausibility. This step marks a reconciliation between computational and biological ontologies: the computational model enables efficiency and prediction, while the biological framework restores empirical validity and legitimacy to the results.

A particularly revealing insight emerges in O6, where AlphaFold's learning process is described as a trajectory of intermediate structures, each representing the model's evolving "belief" about the most probable configuration. Here, the object of study appears emergent and processual rather than static, aligning with contemporary conceptions in data-intensive science, where objects of knowledge are not directly observable but assembled through iterative computation.

Table 4.2 - Key excerpts identified through the ontological lens of the heuristic tool applied to the AlphaFold article. Each excerpt, labeled O1 to O6, is accompanied by an interpretive commentary.

#	Excerpt	Analysis and interpretation
O1	*"The AlphaFold network directly predicts the 3D coordinates of all heavy atoms for a given protein* ***using the primary amino acid sequence*** *and aligned sequences of homologues as inputs"* (page 585)	The object of analysis is clearly defined: the 3D structure of a protein, represented through atomic coordinates. This reflects a positivist stance: the object is measurable and manipulable through computation.
O2	*"The key principle [...] is to view the prediction of protein structures* ***as a graph*** *inference problem in 3D space in which the edges of the graph are defined by residues in proximity"* (page 585)	This formalizes the object as a graph in 3D space, suggesting that the protein is computationally modeled as a network of spatial relationships, abstracting away from its biological materiality.
O3	*"The 3D backbone structure is represented as N_res* ***independent rotations and translations*** *[...]"* (page 586)	The object is treated as a geometrical, manipulable system composed of independent, transformable units.
O4	*"Conversely, the peptide bond* ***geometry is completely unconstrained*** *and the network is observed to frequently violate the chain constraint […]"* (page 586)	During the prediction phase, AlphaFold allows temporary violations of biochemical constraints to prioritize learning and optimization in a flexible computational space. This reflects a computational ontology that treats the object as manipulable, data-driven, and iteratively constructed. However, these violations are not permanent; they reflect a strategy of temporary abstraction rather than ontological indifference to biochemical realism.
O5	*"Exact enforcement of peptide bond geometry is only achieved in the* ***post-prediction relaxation of the structure*** *by gradient descent in the Amber force field."* (page 586)	This post-prediction refinement step reintroduces biochemical realism by aligning the predicted structure with known physical constraints using force field simulation. This does not just 'correct' the model but shows that biological ontology actively constrains and legitimizes the computational object.
O6	*"This provides* ***a trajectory of 192 intermediate structures,*** *in which each intermediate represents the belief of the network of the most likely structure"* (page 587)	The object is not fixed but dynamically constructed: deep learning enables the exploration of multiple possible structures, progressively refined and discarded through weighted optimization. The "trajectory" reveals a processual ontology, where the object emerges through the computational process itself, shaped by the system's learning dynamics and axiological priorities (e.g. accuracy, stereochemical plausibility)

In summary, AlphaFold embodies a hybrid ontological configuration. The protein is conceptualised as a computationally manipulable entity, an object that can be abstracted, transformed, and reconstructed through algorithmic operations, yet always within the boundaries imposed by biological plausibility. The method draws on a computational ontology to achieve its remarkable predictive capacity, while the biological ontology continues to provide the ultimate framework for validating and interpreting the results. This interplay marks a notable redefinition of scientific objecthood: conducting science with AI no longer means merely observing pre-existing entities, but actively constructing and negotiating what counts as a legitimate object of inquiry.

Epistemological assumptions

The epistemological examination of the *AlphaFold* article focuses on how knowledge is produced, validated, and structured within this AI-based scientific method. Based on the excerpts labelled E1–E8 (see Table 4.3), the analysis highlights a distinctive epistemic configuration, spanning the transformation of empirical biological data into computational representations, the central role of large-scale labelled datasets, and the dynamic, iterative character of model-driven knowledge generation.

A first notable aspect is the integration of biological knowledge with statistical and geometric reasoning. Empirical constraints (evolutionary, physical, and spatial) serve as the foundational epistemic sources (E1), yet they are operationalised within a formal and computational framework. For instance, multiple sequence alignments (MSA) function as input data (E3), mediating between biologically grounded information and its statistical reformulation. By interpreting MSA as probabilistic objects, AlphaFold can infer correlations across homologous sequences, uncovering relationships that are not explicitly encoded but gain epistemic significance within the model.

Learning in AlphaFold extends beyond traditional supervised learning approaches. The system employs self-supervised strategies such as self-distillation and self-estimation of accuracy (E2), enabling it to derive internally validated forms of knowledge from unlabeled data. In doing so, AlphaFold generates latent and inferred knowledge that emerges autonomously through its learning dynamics. These processes involve both local and global reasoning, as indicated in E4, where initial low-level representations evolve rapidly to capture multi-scale spatial and evolutionary relationships.

Another central epistemic feature concerns how knowledge is computed and validated. Predictions are not simply derived from data correlations but are structured through optimisation and comparison mechanisms. Validation involves assessing predicted structures against known ones using geometric similarity metrics (E5), reflecting a view of scientific knowledge as spatially precise, quantifiable, and subject to continuous verification.

The role of data in this epistemic configuration is particularly crucial. The model's accuracy depends on the availability of extensive, high-quality labelled datasets. As noted in E8, shallow or incomplete alignments lead to significant drops in predictive performance. When labelled data are scarce, the model generates synthetic labels to augment the training set (E6), effectively transforming raw or incomplete sequences into knowledge-rich resources. This represents a profound epistemological shift: data creation itself becomes part of the scientific method, rather than a preliminary or external step.

Knowledge in AlphaFold is also conceived as evolving and provisional rather than fixed. As described in E7, the model does not immediately converge on a single structural solution but instead

explores a trajectory of possible conformations, each representing a provisional "belief state." Knowledge thus unfolds through an iterative and incremental process, continually refined by both internal optimisation criteria and domain-specific heuristics.

Taken together, these features show that epistemology in AI-based scientific methods becomes deeply intertwined with computational practice, data infrastructure, and machine learning dynamics. Knowledge is not merely discovered but continuously constructed, refined, and stabilised through model-mediated processes, where inference, optimisation, and evaluation operate as central epistemic mechanisms.

Table 4.3 - Key excerpts identified through the epistemological lens of the heuristic tool applied to the AlphaFold article. Each excerpt, labeled *E1* to *E8*, is accompanied by an interpretive commentary.

#	Excerpt	Analysis and interpretation
E1	*"...training procedures based on the **evolutionary, physical and geometric constraints** of protein structures."* (page 584)	Scientific knowledge is grounded in multiple biological constraints: evolutionary, physical, and geometrical.
E2	*"...learning from **unlabelled** protein sequences using self-distillation and self-estimates of accuracy."* (page 585)	Knowledge emerges from unlabelled data via self-supervision and internal accuracy estimation, reflecting an epistemology that includes latent and inferred forms of knowledge.
E3	*"The AlphaFold network directly predicts the 3D coordinates of all heavy atoms for a given protein using the **primary amino acid sequence and aligned sequences** of homologues **as inputs**"* (page 585)	Input knowledge is grounded in empirical biological data (primary sequences and MSA), but MSA are treated as sources of statistical correlations, computationally analyzed to extract relevant evolutionary patterns.
E4	*"Innovations [...] are new mechanisms to exchange information within the MSA and pair representations that enable direct **reasoning about the spatial and evolutionary relationships**. These representations are **initialized in a trivial state** [...]but **rapidly** develop and **refine** [...]. Breaking the chain structure to allow simultaneous **local refinement** of all parts [...]."* (page 585)	Knowledge evolves from an initially trivial state through dynamic reasoning over spatial and evolutionary relationships, involving both global and local refinements.
E5	*"Predictions [...] are **computed** [...]. The final loss compares **predicted** to true positions under many different alignments."* (page 587)	Knowledge is produced computationally and validated through geometric alignment with known data.

E6	*"The AlphaFold architecture is able to train to high accuracy using only supervised learning on PDB data. [...] We use a trained network to predict the structure of around 350,000 diverse sequences from Uniclust3036 and make a* ***new dataset*** *[...]. We then train the same architecture again. [...]."* (page 587)	Self-generated predictions create new labelled data (synthetic and model-generated data), enabling recursive cycles of learning.
E7	*"incremental improvements"; "****trajectories****"* (page 587)	*Knowledge is conceived as an iterative, incremental process, where multiple representations are explored and refined, reflecting a plurality of intermediate, evolving knowledge.*
E8	*"The accuracy decreases substantially when the median alignment depth is less than around 30 sequences."* (page 587-588)	Epistemic reliability depends on the quantity and depth of input data.

Axiological commitments

The axiological analysis of the *AlphaFold* article (based on the textual excerpts presented in Table 4.4), uncovers a set of underlying values that guide the method's design, validation, and epistemic legitimacy. At the core of these values lies a strong emphasis on accuracy (A1), which functions as the principal metric of scientific success. This is expected, given that AlphaFold operates within the paradigm of deep learning, where accuracy serves as a key criterion for model evaluation and optimisation. Every architectural choice, loss function, and training procedure is calibrated to maximise this value.

However, in AlphaFold, accuracy is not interpreted as a purely numerical construct. As indicated in A2, the system's conception of correctness also includes spatial plausibility and orientational realism; predictions must not only reproduce known atomic coordinates but also maintain the geometric coherence characteristic of actual molecular structures. This attention to biologically meaningful configuration introduces a further value of interpretive sense-making, where results must align with established biochemical knowledge even if this limits potential gains in raw accuracy, as illustrated in A4.

Another prominent value is automation (A5). AlphaFold's architecture is explicitly designed to generate its own training data, converting previously unlabeled sequences into labelled datasets. This self-sufficient design reflects a commitment to scalability and efficiency, reducing reliance on human intervention and enabling large-scale, accelerated scientific progress. Through this capacity for autonomous data production and self-improvement, AlphaFold embodies a vision of science in which computation takes on an active and generative epistemic role.

Importantly, this focus on automation does not equate to a reductionist view of scientific practice. As shown in A3 and A7, the model also enacts the value of integration and plurality. Its success depends on the coordination of multiple informational layers (evolutionary, spatial, and structural) and on their

coherent synthesis within a unified framework. Scientific performance thus emerges from the interplay of diverse mechanisms, rather than from any single algorithmic innovation.

Table 4.4 - Key excerpts identified through the axiological lens of the heuristic tool applied to the AlphaFold article. Each excerpt, labeled A1 to A7, is accompanied by an interpretive commentary.

#	Excerpt	Analysis and interpretation
A1	*"AlphaFold greatly improves the **accuracy** of structure prediction…".* *"…to enable accurate end-to-end structure prediction…".* *"…contributes markedly to **accuracy**…".* *"…produce high-accuracy structures…".* (page 584)	Accuracy emerges as the dominant value, consistently used as the core metric of success. Every architectural choice is oriented toward maximizing prediction accuracy, shaping the training process, updates, and evaluation.
A2	*"loss term that places substantial weight on the **orientational correctness** of the residues."* (page 585)	Accuracy is not limited to numerical proximity: spatial plausibility and realistic orientation are valued as key dimensions of model correctness.
A3	*"...ensuring that the overall Evoformer block is able to fully **mix information** [...] and prepare for structure generation..."* (page 586)	The method values coherence and integration across representational levels, where good performance is tied to the capacity to orchestrate consistent and complete information flow.
A4	*"...**final relaxation** does not improve the accuracy... but does **remove distracting** stereochemical violations…"* (page 586-587)	Beyond accuracy, the model integrates a form of sense-making that aligns outputs with pre-existing scientific knowledge, as shown by the attention to stereochemical plausibility.
A5	*"[...] learning from unlabelled protein sequences using self-distillation and self-estimates of accuracy." (page 585)*	Automation emerges as a key value, enabling large-scale structure prediction and automatic generation of labeled data. The model's capacity to self-train and produce new datasets supports an efficient and self-sustaining research pipeline, overcoming time-consuming human intervention and reinforcing scalability and reproducibility.
A6	*"The resulting trajectories are surprisingly smooth after the first few blocks, show- ing that AlphaFold makes constant incremental improvements".* (page 587)	The architecture embodies a principle of continuous refinement, with progress seen as an incremental, smooth trajectory toward higher-quality representations.
A7	*"[...] a variety of different mechanisms contribute to AlphaFold accuracy."* (page 587)	Performance is not attributed to a single mechanism but to a plurality of interacting components, suggesting a complex and layered view of methodological success.

Finally, A6 highlights refinement as a guiding value. AlphaFold's learning trajectory is described as a continuous process of progressive improvement, in which intermediate representations evolve toward increasingly accurate approximations. This reflects a broader epistemic commitment to incrementalism, where scientific advancement is conceived as cumulative, achieved through successive refinements and feedback cycles.

In summary, the axiological layer of AlphaFold reveals a method grounded in the values of accuracy, automation, coherence, and refinement, yet also attuned to the biological plausibility and epistemic depth of the domain it models. These intertwined values illuminate how AI-based scientific methods embed specific visions of what constitutes good, valid, and meaningful science.

4.7 Discussion

The analyses presented in this chapter, both the general mapping of AI-based methods (Section 4.5) and the focused case study of *AlphaFold* (Section 4.6), collectively depict a scientific landscape in transformation. What emerges is not a revolution in the sense of a complete rupture from prior modes of inquiry, but rather a hybrid and progressive evolution, where novelty arises through integration, extension, and recombination of existing methodological traditions.

AI is not merely a new research instrument; it represents an umbrella of techniques grounded in distinct conceptual premises, including data-intensiveness, statistical reasoning, and machine learning's capacity to recognise and generate patterns. These premises expand the horizon of scientific creativity, enabling researchers to ask new types of questions, to approach problems from alternative perspectives, and to reconfigure how inquiry itself is organised. In doing so, AI makes visible (and operational) the long-recognised fact that scientific methods are plural, non-linear, and context-dependent. It does not replace traditional scientific reasoning but renders its underlying complexity explicit.

A closer examination of *Scientific Discovery in the Age of Artificial Intelligence* (Wang et al., 2023) and of the AI-based methods extracted in our analysis reveals a significant shift in the first two scientific functions of method: *data curation* and *data representation*. Doing science today increasingly involves pre-methodological activities, such as decisions about how to select, clean, structure, and represent data before any formal experiment or hypothesis testing can occur. These are epistemic decisions in their own right, shaping what can be known and how it can be known. In this expanded methodological landscape, AI contributes novel techniques such as deep learning, graph-based machine learning, generative models, and natural language processing, alongside specific new procedures for anomaly detection, weak labeling, or data embedding. These are no longer peripheral routines but central components of the epistemic toolkit of contemporary science.

This expanded architecture of methods reflects the ongoing datafication of science. As disciplines become increasingly digital and data volumes grow exponentially, new challenges and opportunities emerge. AI does not merely respond to this context; it actively reshapes what becomes visible, relevant, and inferable. Scientific discovery now proceeds not only through intentional, human-led experimentation, but also through automated, large-scale, and continuous data generation—including simulation and the creation of synthetic datasets. The scientist's role expands to include the design of systems that can detect patterns, model uncertainty, and generate novel hypotheses. In this sense, AI operates as a method of possibility expansion, as clearly illustrated by *AlphaFold*.

The role of representation is pivotal in this shift. When an AI model encodes a molecule as a graph or a protein as a vector in high-dimensional space, it does more than visualise data: it reformulates the ontology of the object itself. The world is re-encoded in forms that are computationally tractable yet remain bound to physical and theoretical constraints. In *AlphaFold*, the protein simultaneously functions as a biological and geometric entity, both meaningful and measurable. This dual representation requires scientists to navigate multiple layers of abstraction, maintaining coherence between computational and biological realism.

The case study also demonstrates the interdependence between epistemic and axiological dimensions. Knowledge is optimised through accuracy, but it is also validated through the value of biological plausibility. The final model must not only fit the data statistically but also make *biological sense*. Scientific understanding emerges at the intersection of precision and interpretation, where a structure converges not just numerically but conceptually.

Equally significant is the context in which knowledge is now produced. The laboratory has expanded beyond physical experiments to include computational environments: data centres, shared repositories, and international collaborative platforms such as CASP. Knowledge emerges from distributed ecosystems, where experimental scientists, computer scientists, domain experts, and AI systems interact. In this new epistemic setting, the algorithm becomes not merely an assistant but a participant in the process of inquiry. Recent philosophical reflections on AI further illuminate the implications of these findings. Floridi (2025) characterises AI as a new form of *artificial agency without intelligence*, capable of enacting rule-based behaviour without possessing human-like understanding. His "Multiple Realisability of Agency" thesis proposes that agency can take multiple forms while preserving its functional core. This perspective aligns with the results of the present analysis, which suggest that methodological rules now function as a shared normative infrastructure across human and AI agents. Both categories of agents enact these rules to interpret data and make methodological decisions, indicating that scientific inquiry is increasingly sustained by a distributed, multi-agent epistemic system. Floridi's conceptualisation thus provides a timely philosophical anchor for interpreting the transformation of scientific methods in the age of AI.

What we are observing, therefore, is the rise of methodological hybridity. Simulation and experimentation no longer stand in opposition; optimisation and interpretation coexist; and hypothesis generation is partially automated yet still demands human judgment and critical reflection. This hybridity is both technical and philosophical, reopening fundamental questions about what constitutes an experiment, what qualifies as an explanation, and what counts as reliable knowledge.

This analysis connects directly with the principle of methodological pluralism discussed in Section 4.4.2. AI does not simply add new methods to the repertoire of science; it introduces an additional dimension of plurality, a meta-level where models not only support reasoning but redefine its contours. It also draws attention to what has traditionally remained invisible in accounts of scientific method: the infrastructural and preparatory work that precedes discovery, the choices about which data to collect, how to preprocess them, and how to make them interpretable to machines. These decisions are epistemically generative, as they shape the very conditions under which hypotheses can be conceived.

In this sense, AI is not merely a tool for testing hypotheses but a method for constructing the space of the thinkable. Its novelty lies not in establishing an entirely new paradigm but in expanding and

rearticulating existing ones, urging a reconsideration of the architecture of scientific inquiry, its timelines, environments, actors (both human and non-human), and values.

For science education, this evolving landscape calls for a shift in how we conceptualise "the scientific method". Rather than presenting it as a fixed sequence of steps, we must understand it as a dynamic, context-sensitive rationality, one increasingly shaped by data, models, and computation. Recognising this transformation is essential for preparing future generations of scientists and educators to engage with a world where the boundaries between experimentation, simulation, and algorithmic reasoning are ever more intertwined.

4.8 Conclusions

This chapter has examined how AI is reshaping the methodological foundations of scientific inquiry. By mapping AI-based methods and analysing the *AlphaFold* case through the heuristic tool, it has shown that AI introduces not a rupture but an expansion of methodological reasoning, where new computational practices intertwine with traditional scientific values and norms. Across the ontological, epistemological, and axiological dimensions, AI emerges as both a tool and an agent of knowledge production. Moreover, the scientific object in the case we examined (i.e. protein) emerges as a dynamic interplay between computational and biological entities.

These findings suggest that scientific inquiry is evolving into a hybrid, distributed enterprise, where humans and machines collaboratively enact methodological rules and co-construct meaning. Understanding this transformation is essential not only for philosophy of science but also for science education, where teaching about methods must now reflect this emerging reality. Building on the validity and analytical power demonstrated by the heuristic tool in this analysis, the next chapter applies it again to examine the methodological assumptions underpinning specific cases in science education research, thereby extending the reflection from scientific inquiry to its educational counterpart.

References

Allchin, D. (2017). Beyond the Consensus View: Whole Science. *Canadian Journal of Science, Mathematics and Technology Education*, 17(1), 18–26. https://doi.org/10.1080/14926156.2016.1271921

Barelli, E., Lodi, M., Branchetti, L., & Levrini, O. (2024) Epistemic Insights as Design Principles for a Teaching-Learning Module on Artificial Intelligence. *Science & Education.* https://doi.org/10.1007/s11191-024-00504-4

Billingsley, B., Grzes, M., & Annetta, L. (2025). Epistemic insight and artificial intelligence [Editorial]. *Science & Education, 34*(2), 239–244. https://doi.org/10.1007/s11191-024-00448-2

Brandon, R.N. (1994). Theory and experiment in evolutionary biology. *Synthese* 99, 59–73. https://doi.org/10.1007/BF01064530

Caramaschi, M., Levrini, O., & Erduran, S. (submitted). Learning from AlphaFold: AI-Driven Methodological Shift in Science for Science Education.

Chen, B., Xu, Y., Liu, H. *et al.* (2024). Comparing Practical Items in High-Stake Exams in Different Science Subjects: in View of the Diversity of Scientific Methods. *Sci & Educ* **33**, 1419–1434. https://doi.org/10.1007/s11191-023-00437-4

Cooper, G. (2023). Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence. *Journal of Science Education and Technology*. 32 (*1*), 444–452. https://doi.org/10.1007/s10956-023-10039-y

Cullinane, A., Erduran, S., & Wooding, S.J. (2019). Investigating the diversity of scientific methods in highstakes chemistry examinations in England. *International Journal of Science Education*, 41(16), 2201–2217. https://doi.org/10.1080/09500693.2019.1666216

Dill, K. A., Ozkan, S. B., Shell, M. S. & Weikl, T. R. (2008). The protein folding problem. *Annu. Rev. Biophys.* 37, 289–316.

Ding, L. (2019). Theoretical perspectives of quantitative physics education research. *Physical Review Physics Education Research*, 15, 020101. doi: 10.1103/PhysRevPhysEducRes.15.020101

Erduran, S. (2023). AI is transforming how science is done. Science education must reflect this change. *Science,* 382, 9788. https://doi.org/10.1126/science.adm9788

Erduran, S., & Dagher, Z. R. (2014). Reconceptualizing the Nature of Science for Science Education. Scientific Knowledge, Practices and Other Family Categories. Springer.

Erduran, S. & Levrini, O. (2024) The impact of artificial intelligence on scientific practices: an emergent area of research for science education. *International Journal of Science Education*, 46(18), 1982–1989. https://doi.org/10.1080/09500693.2024.2306604

Erduran, S., Simon, S. and Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's Argument Pattern for studying science discourse. Sci. Ed., 88: 915-933. https://doi.org/10.1002/sce.20012

Floridi, L. (2025). AI as Agency without Intelligence: On Artificial Intelligence as a New Form of Artificial Agency and the Multiple Realisability of Agency Thesis. Philos. Technol. 38, 30. https://doi.org/10.1007/s13347-025-00858-9

Frodeman, R. (1995). Geological reasoning: Geology as an interpretive and historical science. Geological Society of America Bulletin, 107, 960-968.

Ioannidou, O., & Erduran, S. (2021). Beyond Hypothesis Testing. *Science & Education* 30, 345–364. https://doi.org/10.1007/s11191-020-00185-9

Irzik, G., & Nola, R. (2011). A family resemblance approach to the nature of science for science education. *Science & Education*, 20, 591-607.

Irzik, G., & Nola, R., (2014). New directions for nature of science research. International Handbook of Research in History, Philosophy and Science Teaching, (Ed.) M. Matthews. Springer, 999-1022.

Jia, F., Sun, D., & Looi, C. K. (2024). Artificial Intelligence in Science Education (2013–2023): Research Trends in Ten Years. *Journal of Science Education and Technology,* 33, 94–117. https://doi.org/10.1007/s10956-023-10077-6

Jumper, J., Evans, R., Pritzel, A. *et al.* (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2

Lederman, N.G., Abd-El-Khalick, F., Bell, R.L., & Schwartz, R.S. (2002). Views of Nature of Science Questionnaire: Toward Valid and Meaningful Assessment of Learners' Conceptions of Nature of Science. *Journal of Research in Science Teaching*, VOL. 39, NO. 6, 497-521.

Liu, H., Chen, B., Ma, J. *et al.* (2024). An Investigation of High School Students' Attitudes and Perceptions About the Diversity of Scientific Methods in Chemistry Learning. *Sci & Educ*. https://doi.org/10.1007/s11191-024-00589-x

National Science Teachers Association. (1982). Science-technology-society: Science education for the 1980s (An NSTA position statement). Washington, DC: Author.

Osborne, J., Collins, S., Ratcliffe, M., Millar, R., & Duschl, R. (2003). What "Ideas-about-Science" should be taught in school science? A Delphi study of the expert community. *Journal of Research in Science Education*, 40(7), 692-720.

Senior, A. W. et al. (2019). Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13). Proteins 87, 1141–1148.

Upahi, J. E., Ramnarain, U., Alabi, T. P., & Jimoh, M. A. (2025). Examining the diversity of scientific methods in the South African National Senior Certificate chemistry papers. *Research in Science & Technological Education*, 1–18. https://doi.org/10.1080/02635143.2025.2557931

Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., Chandak, P., Liu, S., Van Katwyk, P., Deac, A., Anandkumar, A., Bergen, K., Gomes, C. P., Ho, S., Kohli, P., Lasenby, J., Leskovec, J., Liu, T.-Y., Manrai, A., … Zitnik, M. (2023). Scientific discovery in the age of artificial intelligence. *Nature* 620, 47–60. https://doi.org/10.1038/s41586-023-06221-2

Wivagg, D., & Allchin, D. (2002). The Dogma of "The" Scientific Method. *The American Biology Teacher*. 64, 645-646. https://doi.org/10.2307/4451400

Woodcock, B. A. (2014). "The Scientific Method" as myth and ideal. *Science & Education*, 23: 2069-2093.

Wulff, P., Kubsch, M., & Krist, C. (Eds.). (2025). Applying machine learning in science education research: When, how, and why? Springer. https://link.springer.com/book/10.1007/978-3-031-74227-9

Yeh, Y., Erduran, S., Hsu, Y. (2019). Investigating Coherence About Nature of Science in Science Curriculum Documents. Taiwan as a Case Study. *Science & Education*, 28, 291-310. https://doi.org/10.1007/s11191-019-00053-1

Chapter 5 – AI and qualitative methods in dialogue

Toward a re-discussion of mixed methods in Science Education Research

5.1 Introduction

Having examined how AI-based methods are transforming research practices in science more generally, this chapter returns to the field of science education to explore how such methods are currently being adopted, combined, and interpreted in educational research. At this stage, both the heuristic tool and the understanding of AI-based approaches are mature enough to be, respectively, applied to and explored in recent empirical work in the field. This chapter thus combines empirical analysis of exemplar cases with conceptual discussion of the emerging features that may reconfigure the notion of mixed methods in science education (Section 5.4).

Specifically, three recent and exemplary studies in science education are presented as they employ computational techniques (machine learning, network analysis, and NLP) to analyze textual data. For each study, the principal publication was analysed and an expert interview with one of the lead researchers was also conducted, both supported by the use of the heuristic tool (Sections 5.5 and 5.6). This chapter reflects on the methodological and epistemological tensions involved in merging qualitative and computational approaches. Moreover, it traces signs of change around mixed methods in educational research to better understand how it is evolving (Section 5.7). Then, the chapter concludes by proposing new directions for method integration and re-opens the debate on how mixed methods may be redefined in the age of AI (Section 5.7.2).

5.2 Conceptual background and main statements

As discussed in Chapter 1, the methodological system of science education research is currently undergoing significant transformations. Digitalisation, the increasing computational capacity of computers, and the introduction of artificial intelligence techniques have opened new possibilities in research practices. As shown in Chapter 2, the available methods and the opportunities for combining them are expanding. These changes are not limited to technical tools but also involve novel approaches in which humans and machines (data-driven models) develop knowledge together. In some cases, existing paradigms appear insufficient to capture situations that no longer fully align with their idealised forms. For these reasons, a heuristic tool and related metrics were developed in Chapter 3 to analyse and map these ongoing transformations more systematically.

In this chapter, we focus on a distinctive set of research studies in the field of science education, centred on the analysis of textual data and emerging in the last five years. Their novelty lies, for example, in the explicit adoption of computational and AI-based techniques in qualitative studies, such as the first case study conducted in Chapter 2 that integrated natural language processing with a thematic literature review. In such studies, the computational contribution (for instance, embeddings, clustering, or network modelling) is not merely an ancillary technical step but becomes dialogically entangled with the qualitative contribution of the human researcher. These studies typically start from data that are predominantly qualitative in nature, which are then transformed into a second, computationally tractable representation.

This leads to the central argument of this chapter: we argue that there are important novelties concerning three conceptual structures, previously introduced in Chapter 1. We briefly recall them here as a reference horizon for the analyses that follow:

- **Mixed methods as a research paradigm:** following Johnson and Onwuegbuzie (2004), mixed methods have been proposed as a "third paradigm" alongside qualitative and quantitative traditions. Within this framework, qualitative and quantitative studies are pragmatically combined to enrich the understanding of educational phenomena.
- **Types of mixed methods designs:** four main designs describe how qualitative and quantitative studies can be integrated, from sequential to concurrent and embedded approaches. These designs are typically procedural in nature, focusing on when and how data are collected and combined.
- **The four paradigms of science education research:** as outlined by Treagust and Won, science education research can be mapped into four paradigms: post-positivism, interpretivism, mixed methods, and critical theory. Each paradigm is associated with specific research topics, preferred designs, roles of the researcher, standards of quality, and styles of reporting.

Through the analysis of selected case studies in the literature, the following sections will revisit these foundational concepts. In particular, the core statements advanced are that (1) mixed methods are not merely driven by pragmatic considerations; (2) cases where computational and AI-based methods are integrated with qualitative analysis represent a second generation of mixed methods; and (3) the mixed-methods paradigm is extending and reconfiguring its scope by incorporating a third pillar alongside qualitative and quantitative approaches, that is the computational approach.

This study presents evidence to show how such integrations operate and what assumptions they embody. At the same time, the broader debate remains open. Our aim is therefore to contribute to re-opening the discussion on methodological foundations within the broader research community, recognising that methodological reflection is itself a collective and evolving endeavour.

5.3 Aims and research questions

The aim of this chapter is to investigate how computational methods are being combined with qualitative approaches for text data analysis in science education research. The chapter further aims to explore whether such combinations can be meaningfully understood within the framework of mixed methods.

Accordingly, this chapter addresses the following research questions:

1. What studies combining qualitative and computational approaches are emerging in science education research for textual data analysis, and what methodological features do they display?
2. Is it possible to speak of a "new genre" of mixed methods? If so, what novelty do they display compared to previous approaches?
3. Do these forms of research represent a change in research paradigms, for example, an evolution of the existing ones, or the emergence of a new paradigm?

Together, these questions invite a critical examination of how methodological practices are shifting when computational tools are integrated into the analysis of educational data, and how such shifts

might reshape the conceptualization of mixed methods, a notion has remained relatively crystallised over the past two decades and less frequently explored in recent years.

By situating the inquiry at the intersection of practice (often visible, practical, and explicit) and the foundational assumptions of methodology (frequently overlooked and implicit), the chapter aims to foster a deeper reflection on the use and integration of computational or AI methods with qualitative ones in the current landscape of science education research. At the same time, it seeks to articulate the methodological transformations that are underway, and to consider how these may contribute to redefining research paradigms.

5.4 Methodology

In order to address the research questions outlined in Section 5.3, this study adopted a twofold methodological strategy: the analysis of exemplar studies through the heuristic tool, and semi-structured interviews with their lead authors.

Therefore, we initially selected a set of exemplar studies from recent science education research that combine qualitative and computational approaches to the analysis of textual data. Specifically, we refer to computational approaches that come from the field of AI, as described in Chapter 2.3. These selected studies were analyzed through the heuristic tool developed in Chapter 3, which was operationalized to surface the ontological, epistemological, and axiological assumptions underpinning methodological choices. Second, in order to complement the document analysis with researchers' own perspectives, we conducted expert interviews with the lead authors of the selected studies. The interviews were designed around the same heuristic dimensions, thus enabling a dialogical exploration of how these methods are perceived, justified, and enacted in practice. Below, we present further methodological details.

5.4.1 Case selection

The selected studies were not intended to constitute a comprehensive review of all science education research employing computational approaches, but rather to provide illustrative examples of emerging methodological integrations. The criteria for selection were:

A. Explicit use of computational or AI-based methods for textual data analysis;
B. Methodological novelty in combining computational and qualitative approaches;
C. Diversity of methods across the examples;
D. Relevance to ongoing debates on mixed methods in educational research.

Based on these criteria, three studies were chosen as primary cases, focusing respectively on Machine Learning with natural language processing, AI into a Computational Grounded Theory approach, and Thematic Network Analysis (Wulff et al., 2023; Mariegaard et al., 2022; Tschisgale et al., 2023; Wulff et al., 2022). In addition, insights from the Latent Dirichlet Allocation method, previously implemented and discussed in Section 2.4, were used to inform and enrich the interpretation of results, providing a point of comparison grounded in prior practical experience.

5.4.2 Operationalising the heuristic tool and literature analysis

The heuristic tool presented in Chapter 3 was employed to guide the analysis of both the selected articles and the expert interviews. The tool was operationalized in the following way:

1. Identification of relevant passages: each article was read with the support of the guiding questions of the tool, which served as markers to identify explicit and implicit statements concerning ontology, epistemology, and axiology.
2. Coding and annotation: extracted passages were annotated according to the tool's categories and markers (e.g., object of analysis, type of data, role of the researcher, values embedded in knowledge).
3. Mapping: the coded information was organized into a comparative grid, which enabled a systematic visualization of the foundational assumptions underlying each method.
4. Triangulation: Coding results were discussed among three experts in order to enhance reliability and reduce individual bias
5. Synthesis: the resulting maps were used to compare across cases, highlighting both continuities with established traditions and novelties that suggest possible new directions for mixed methods.

5.4.3 Interview protocol design for expert interviews

To complement the document analysis with the perspectives of the researchers themselves, semi-structured interviews were conducted with the lead authors of the selected studies. The aim was not only to validate the insights emerging from the article analysis, but also to open a dialogical space in which authors could articulate the implicit rationales, assumptions, and methodological choices that are often condensed or hidden in published work. In this way, the interviews served both as a confirmatory and an exploratory instrument, enriching the analytical framework with lived experiences and reflective accounts.

The interview protocol was designed following the Interview Protocol Refinement (IPR) framework proposed by Castillo-Montoya (2016). This framework is particularly suited for qualitative research, as it ensures the reliability and richness of data by structuring the development of the protocol through four iterative phases:

1. **Aligning questions with the research questions**, to ensure coherence between the interview and the overall study.
2. **Constructing an inquiry-based conversation**, to foster a dialogical atmosphere, avoiding excessive detachment or ambiguity.
3. **Receiving feedback on the interview protocol**, to refine clarity, sequence, and alignment with the study's aims.
4. **Piloting the interview protocol**, to verify the effectiveness of the questions and flow in practice.

In the first phase, I selected interview questions that did not mirror the research questions directly but were instead designed to elicit information that could address them indirectly. This distinction was crucial: while the research questions guided the overall investigation, the interview questions functioned as tools to open up reflections on ontological, epistemological, and axiological

assumptions, as well as on methodological implications and the positioning of computational-qualitative methods within mixed-methods research.

The second phase involved shaping the tone and progression of the dialogue. Particular care was taken to use language that invited researchers to share their experience without creating unnecessary distance, while also preventing misunderstandings. Following Castillo-Montoya's recommendation, the protocol moved from short and descriptive questions (e.g., academic background, familiarity with the method) toward progressively more focused inquiries about their study, and finally to broader reflective questions concerning the paradigmatic status of computational-qualitative methods. This sequencing allowed participants to ease into the conversation and culminate in open reflections on their epistemological and methodological positioning.

The third phase of feedback and refinement was conducted in collaboration with my supervisor. Together we reviewed the wording, ordering, and alignment of questions to ensure that the interviews could generate material relevant to the study while remaining conversational and natural.

Finally, the piloting phase was conducted during the first interview. Although the estimated duration of one hour was exceeded (lasting approximately 75 minutes), this pilot confirmed that all questions were meaningful, well-constructed, and appropriately sequenced. No substantial revisions were required, apart from recalibrating the expected interview length.

The resulting protocol is presented in Table 5.1, where each question is organized by type (introductory, transition, specific, and general), and linked to the kind of information it was intended to elicit. The final structure reflects the dual orientation of the study: on the one hand, an analytical focus on the ontological, epistemological, and axiological dimensions embedded in methodological choices; on the other hand, a dialogical and exploratory inquiry into whether, and in what sense, approaches such as these selected can be considered part of a "new genre" of mixed methods in Science Education Research.

Table 5.1 - Interview protocol

Type of questions	**Questions**	**Information acquired**
Informative questions	Q1) What is your academic background? Q2) [BRIEFLY] What are your main research interests, and how would you describe the field you work in? Q3) How long have you been working as a researcher in this area? Q4) How would you describe your familiarity with methods in your own research practice?	- Understanding the kind of expertise as researcher in SER and on methods

Transition questions	Q5) When and how did you first come into contact with the method called *Computational Grounded Theory/NLP and ML/Thematic Network Analysis*? Q6) Is this a method you've used in multiple studies, or was it applied specifically in the study titled *[insert article title]*?	- Understanding the confidence with *CGT/NLP and ML/TNA*
Specific questions	Q7) What were your motivations for adopting *CGT/NLP and ML/TNA* in this particular study? What were you hoping it would enable you to do? Q8) In the application of *CGT/NLP and ML/TNA* described in your article, at which points in the process did the human researcher play a particularly active role, and where did the AI take the lead? How would you describe the interaction between these two contributions? Q9) Would you describe the interaction between human and machine as more sequential or dialogical? Did one shape the other? Q10) Do you consider the patterns identified through *CGT/NLP and ML/TNA* as already existing in the data, or do you see them as being constructed through the method? Q11) How did theoretical concepts enter the research process? Were they predefined, or did they emerge from the data? Q12) Do you think *CGT/NLP and ML/TNA* is particularly suited for specific kinds of research questions or phenomena? If so, which ones?	- Specific knowledge on the mixed method under investigation (*CGT/NLP and ML/TNA)* - Axiology: Values of the research and what values CGT/NLP and ML/TNA embed - Epistemology: Exploring the integration of different knowledge - Ontology: what reality *CGT/NLP and ML/TNA* is suited to analyse; the pattern are pre-existing or not - Role of theory

General questions	Q13) Do you consider *CGT/NLP and ML/TNA* a mixed-methods approach? Why or why not? Q14) What are the main epistemological or methodological challenges in combining qualitative and computational approaches like this? Q15) What kinds of skills, competencies, or resources do you think future researchers need to be able to use *CGT/NLP and ML/TNA* effectively? Q16) Do you think using *CGT/NLP and ML/TNA* pushes us to rethink traditional boundaries between qualitative and quantitative research? Q17) Finally, is there anything about your use of *CGT/NLP and ML/TNA* that surprised you, either in terms of challenges, insights, or outcomes?	- collecting "words" to describe this method as mixed methods - epistemological limits of *CGT/NLP and ML/TNA* - future competences in SER related to AI-qual mixed methods

5.4.4 Interviews' process of analysis

The analysis of the interviews was conducted in two complementary phases, combining a technical and a reflexive approach. In the first phase, I worked from the notes taken during the interviews, which were organised by question. These notes served as an initial layer of attention, highlighting the elements that appeared most meaningful during the conversation. Subsequently, I carried out a close reading of the full interview transcripts, aiming to reconstruct each participant's reasoning as a coherent narrative about their research practice. Because all interviews followed a common protocol, similar discursive patterns emerged across cases, allowing for a structured comparison of the researchers' perspectives.

From this iterative process, five recurrent dimensions were identified, which became the framework for the subsequent interpretation):

a) the researcher's background and motivations for adopting the method;
b) the roles attributed to the human analyst, the computational tool, and the theoretical framework;
c) the epistemic, ontological, and axiological assumptions underlying their approach;
d) their positioning with respect to mixed-methods research; and
e) the envisioned applications, competences, and potentials of the method within the Science Education community.

These five dimensions mirror the internal logic of the interviews themselves. The first (a) captures the "past" dimension of methodological motivation, the personal and disciplinary drivers that led researchers to integrate computational tools into qualitative analyses. Across all interviews, a common aspiration emerged: the intention to improve and expand traditional qualitative analysis by harnessing the potential of computational and AI-based methods. For each of them, AI was conceived not as a replacement for qualitative inquiry but as its *extension*, an instrument for establishing a productive dialogue between the computational model and the human researcher. These approaches enabled them

to keep “a foot” in both worlds: the interpretive depth of qualitative research and the systematic, scalable capacities of computation.

The last dimension (e) projects toward the “future”, describing how interviewees envision the further development of such approaches, the competences required, and their potential value for the SER community. The central dimensions (b–d) represent the “present” analytical practice and foundations for applying these approaches with awareness and care. I describe the distribution of roles among human, theoretical, and computational agents; the implicit assumptions and values informing the research; and the explicit positioning of these methods within the broader mixed-methods landscape. Together, these five components form a temporal and epistemic map of the researchers’ methodological reasoning.

The analytical work was not a mechanical process of coding but a reflexive interpretation guided by the heuristic framework developed in Chapter 3, which allowed me to keep my attention toward the guiding questions and markers. Each interview is described following the five-parts articulation, to trace each participant's case and voice (see the structure of the interview description in Fig. 5.1). Then the cases were compared to identify convergences and distinctive elements. In this sense, the procedure paralleled the document analysis conducted earlier, but added a dialogical layer that brought to light the lived reasoning behind methodological choices.

Interviewees and ethical considerations

The interviews conducted for this study involved researchers who had developed or applied the methods in the three selected cases. Each of them was invited not as a private individual but as a *knowledge actor*, a scholar whose methodological choices and reflections have shaped current research practices in Science Education. In this sense, their contributions represent an integral part of the epistemic landscape that this thesis aims to explore.

To respect academic integrity and personal privacy, I have refrained from focusing on biographical or personal aspects that might identify the interviewees beyond what is already evident from their published work. However, because each interview refers to publicly available studies authored by small teams, complete anonymity would be neither possible nor desirable. In this context, naming the interviewees serves not to individualise but to acknowledge their intellectual contribution, and to value the openness with which they have discussed their methodological reasoning. By presenting their voices, my intention is to recognise their contribution to the collective methodological reflection in our field, and to invite the community to engage with these ideas in the same spirit of openness and collaboration in which they were offered.

The results of this interpretive process are presented in the following sections, each organised according to the five analytical dimensions outlined above. This structure enables both within-case coherence and cross-case comparability, supporting the synthesis developed later in this chapter, where the interviews are re-examined alongside the article analyses to draw broader methodological conclusions.

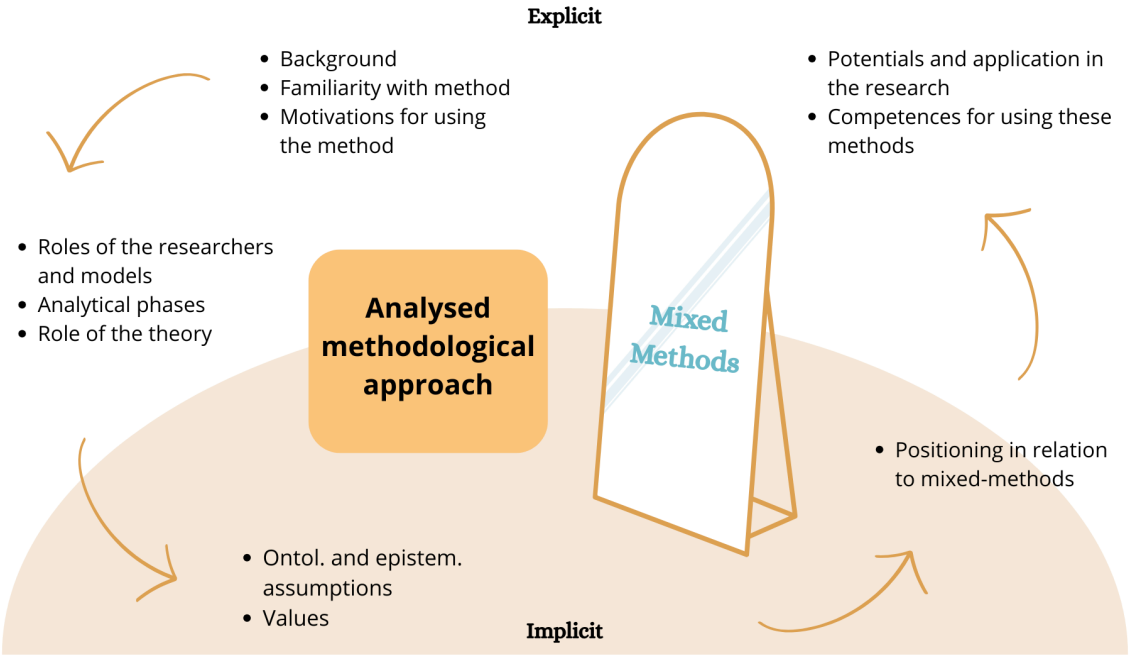

Fig. 5.1 - Each interview is described following the same structure, from background and motivation of implementing computational AI methods with qualitative ones, to the role of researcher, model and theory within the analytical process, to the assumptions embedded in the methodology, then the perspective of the interviewee to call the method as mixed, and finally future potentials and competences.

5.5 Results: mapping methodological assumptions in exemplar cases

In this section we present the results of our analysis of the ontological, epistemological, and axiological assumptions underlying the three selected case studies. We highlight the foundations of each methodology, drawing on significant sentences from the articles that explicitly or implicitly define what is considered the object of analysis, how this object is understood as knowable and what counts as valid knowledge, as well as the values that guide the research. In several cases, the markers of our heuristic tool were useful and crucial in identifying passages where assumptions were not directly stated but clearly at play.

Although the results are presented by distinguishing the three analytical lenses (i.e. ontological assumptions, epistemological commitments, and axiological commitments) this separation should be understood as primarily analytical. In practice, some excerpts may inform more than one dimension at the same time (for example, revealing both ontological and epistemological assumptions). In fact, as discussed in Chapter 3, these foundational dimensions are deeply interrelated: methodological assumptions influence and shape one another, and forcing a rigid separation among them would obscure the coherence and interdependence that characterise each methodological dimension. For these reasons, at times more than one assumption will be presented together within a section that is formally dedicated to a single analytical dimension.

5.5.1 Case 1: Machine Learning and NLP into qualitative analysis

The first case is the article *"Enhancing writing analytics in science education research with machine learning and natural language processing. Formative assessment of science and non-science*

preservice teachers' written reflections" by Wulff et al. (2022). In this study, the authors explore the use of machine learning (ML) and natural language processing (NLP) to analyze reflective writing by preservice teachers, both from physics (domain experts) and non-physics backgrounds (domain novices).

The method involves two main stages: (1) filtering text segments that contain higher-level reasoning elements (such as evaluations, alternatives, or consequences), using a pretrained ML model that was further fine-tuned with the study's data; and (2) clustering these segments through unsupervised ML algorithms applied to language model embeddings (e.g., BERT) in order to identify themes and forms of knowledge represented in the reflections.

The tools include the BERT language model for generating embeddings, UMAP for dimensionality reduction, and HDBSCAN for clustering. The analysis is anchored in a qualitative theoretical model, that is the reflection-supporting model, which distinguishes categories such as circumstances, description, evaluation, alternatives, and consequences.

This study is a relevant example of mixed methods because it combines computational techniques (ML and NLP for segmentation and clustering) with qualitative approaches (theoretical models of reflection and educational interpretation of text). In doing so, it represents an advancement over traditional manual qualitative analysis, addressing issues of scalability and reliability.

Ontological assumptions

In this study, the phenomena under investigation are the ways preservice teachers describe a situation and the reasoning processes they articulate in their reflections. From an ontological perspective, two key assumptions emerge.

First, it is assumed that these descriptions are specific, detailed, nuanced. Thus, there is an assumption on the nature of written texts as not merely containers of answers but carriers of complex knowledge and cognitive processes:

"Writing assignments are typically scored with holistic, summative coding rubrics. This, however, is not very responsive to the more fine-grained features of text composition and represented knowledge in texts."

This reflects an interpretivist orientation in which language is considered a window into the richness of thought and experience, and must be understood in relation to the context in which it is produced.

At the same time, texts are assumed to contain identifiable structures, patterns, and regularities that mirror differences in the reality under study. These become visible when sufficiently large and comparable samples are analyzed:

"ML and NLP are used to filter higher-level reasoning sentences in physics and non-physics teachers' written reflections on a standardized teaching vignette."

Here, the reasoning processes of teachers are assumed to differ structurally depending on whether they are experts in the field or not. Such cognitive structures are understood as manifesting in specific linguistic segments, and can be *filtered*.

Thus, the reality under study is conceived as dual: on the one hand, nuanced and open to interpretation, as in classical qualitative studies; on the other hand, characterized by recurring structures and regularities that reflect differences across groups. These patterned features are not mere linguistic artifacts, but traces of underlying reasoning processes and ways of thinking that emerge from teachers' professional backgrounds and experiences. In this view, language is both a medium that conveys meaning and experience, and a repository of cognitive categories and reasoning modes that can, in principle, be made visible through systematic inquiry.

Epistemological commitments

At the epistemological level, this study incorporates a wide range of choices and assumptions that shape how knowledge is produced. In what follows, we make these commitments explicit by examining and discussing significant sentences from the article across different dimensions of knowledge production.

As already noted, we are considering cases where textual data are analysed within the field of SER. In this instance, the data consist of written reflections on a standardized teaching vignette, produced by both physics and non-physics preservice teachers. The type of knowledge the study seeks to generate is an "analytic, formative assessment", which differs substantially from traditional summative assessment because it is richer, more nuanced, and more responsive to textual features.

Knowledge about these reflections is rendered knowable through a combination of ML and NLP techniques. Sentences are automatically filtered to isolate those containing higher-level reasoning, and then clustered to reveal themes, categories of knowledge, and coherence structures. As the authors explain:

"The filtered sentences are then clustered with ML and NLP to identify themes and represented knowledge in the teachers' written reflections."

This procedure rests on a key assumption: text can be analysed computationally. In other words, natural language can be transformed into alternative representations that preserve semantic features while making them numerical and quantifiable. The method relies, in particular, on computationally transformed textual data such as sentence embeddings derived from pretrained language models (e.g. BERT).

Here lies a crucial epistemological commitment: the belief that metrics within algorithms can capture features that correspond to properties of the real text. In other words, it is assumed that one can observe, in computational results such as coherence indicators or cluster structures, meaningful patterns of reasoning embedded in language. This assumption is made explicit, for example, when the authors note that:

"Consequently, we expected domain experts' texts to have a higher degree of coherence between the sentences... A coherence indicator was calculated on the basis of the contextualized segment embeddings via sentence transformers."

Yet, the authors are careful to ensure that these computational indicators are not detached from human interpretation, which remains central to the analytical process. In fact, knowledge manipulation and production are consistently placed in dialogue with human-coded categories and theoretical models of reflective practice. Validation occurs through comparisons between machine-generated and

human-generated labels: *"To externally validate the quality indicators… Agreement between human labels ('high' versus 'low') with automatically determined labels was calculated."* Moreover, theoretical taxonomies are explicitly introduced as a top-down frame for interpretation, providing domain-specific knowledge that grounds the computational outputs.

The epistemological stance underpinning the study is therefore that machine learning can enhance the production of valid educational knowledge by enabling analytic, formative assessment that transcends the limitations of holistic rubrics. By computationally modelling aspects of teachers' reasoning processes, ML and NLP are positioned as methodological partners that can generate new, more nuanced understandings of reflective writing, provided that their results remain anchored to theoretical and human interpretative frameworks.

Axiological commitments

The values guiding the study are explicitly educational and methodological. On the one hand, there is a strong emphasis on supporting teacher education, by developing means to capture the quality of reflective writing and to differentiate between domain experts and novices. As the authors note:

"Writing assignments are typically scored with holistic, summative coding rubrics. This, however, is not very responsive to the more fine-grained features of text composition and represented knowledge in texts"

In this sentence, they stress the need for a more detailed and sophisticated comprehension of textual knowledge, which motivates the use of computational tools to enhance writing analytics.

On the other hand, the study foregrounds values such as scalability, validity, and reliability, since the proposed methodology is designed to overcome the limitations of traditional manual coding and purely qualitative analysis. As the authors observe:

"While ML and NLP cannot necessarily resolve the challenges around high-inferential coding categories and ambiguity in teachers' language… they might well facilitate the implementation of reliable and valid coding for well-defined (sub-)tasks."

Computational methods are therefore valued both for their ability to reduce the resource intensity of qualitative coding and for their potential to address the bluntness of conventional statistical approaches, which often fail to capture subtle, complex features of language. This limitation is highlighted when the authors write that: *"Commonly used quantitative statistical methods such as parametric, linear models… are mostly incapable to model the complex relationships that are characteristic for language artifacts."*

Finally, the authors recognize the importance of ethical and social values, especially in relation to the potential biases and opacity of machine learning systems. They explicitly acknowledge that:

"Pretrained language models are commonly trained on language sources such as Wikipedia and the Internet where all sorts of gender and racial biases in language are present… Our models would have to be examined with regards to gender biases or similar biases."

This reflection demonstrates that methodological innovation is not only technical, but also normatively charged, as it requires an awareness of the values embedded in computational systems and their social implications.

5.5.2 Case 2: Computational Grounded Theory

The second case is the article *"Integrating artificial intelligence-based methods into qualitative research in physics education research: A case for computational grounded theory"* by Tschisgale, Wulff, and Kubsch (2023). This paper is particularly significant because it applies, for the first time in science education research, the method of Computational Grounded Theory (CGT). This approach represents an explicit attempt to bridge the tradition of qualitative grounded theory with the possibilities opened by artificial intelligence, specifically natural language processing (NLP) and machine learning (ML).

Grounded Theory (GT) has long been one of the most widely used qualitative methods in SER for developing data-driven theoretical insights (Glaser and Strauss, 1967; Vollstedt & Rezat, 2019). It relies on the researcher's analytical sensitivity and interpretive ability to identify meaningful codes, progressively group them into categories, and generate an emergent theory grounded in the data. Classical GT unfolds through iterative cycles of *open coding*, *axial coding*, and *theoretical integration*, aiming to construct a "humble theory" that captures recurrent features of a phenomenon without detaching from empirical evidence.

Computational Grounded Theory was conceived in parallel with this qualitative tradition but introduces a computational layer to support the detection, refinement, and confirmation of patterns. The core idea is to combine human and machine capabilities in a complementary process: the algorithmic procedures enhance scalability and reproducibility, while human interpretive judgment ensures theoretical coherence and contextual sensitivity. The approach remains fully data-driven, yet it redefines the way inductive reasoning unfolds in practice by introducing computational operations as epistemic partners in the analytical process.

In operational terms, the CGT process described by Tschisgale et al. (2023) unfolds through three main phases (Figure 5.2):

- **Pattern detection:** texts are segmented into sentences, transformed into numerical embeddings through pretrained language models, and clustered with unsupervised ML algorithms to detect recurring patterns.
- **Pattern refinement:** the clusters are interpreted through "computationally guided deep reading". Researchers read representative excerpts, develop categories, and refine them in dialogue with substantive theory—in this case focusing on students' approaches to physics problem solving.
- **Pattern confirmation:** the robustness of the identified themes is tested using supervised ML techniques (e.g., relevance vector machines), ensuring that the refined patterns hold across the full dataset and can serve as a basis for automatic coding.

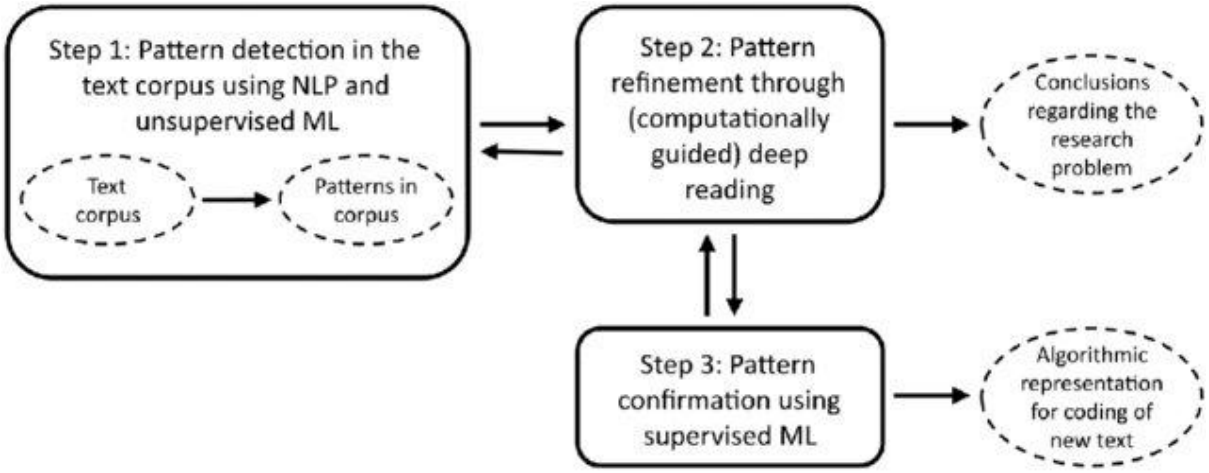

Figure 5.2 - Schematic representation of the three steps of CGT from Tschisgale et al. (2023)

The article also applies CGT in a concrete research context. The study analysed 417 textual descriptions from students, both Physics Olympiad participants and non-participants, about their approaches to problem solving. Tools included pretrained German language models, UMAP for dimensionality reduction, HDBSCAN for clustering, and supervised validation techniques.

This case exemplifies new forms of mixed methods because it systematically integrates qualitative traditions (inductive coding and theoretical interpretation) with computational tools (ML and NLP for pattern detection and validation). It advances grounded theory by addressing its traditional limitations of reproducibility and scalability, while preserving its core inductive and theory-generative logic.

Ontological assumptions

The article describes how qualitative methods (such as GT), which are central in science education research, are used to derive knowledge from data analysis:

"Qualitative research methods have provided key insights […] they share two essential steps: recognizing patterns in the data and interpreting these patterns."

In this excerpt, the authors refer to qualitative analysis, which often relies primarily on textual data. From this, it is possible to infer certain epistemological and ontological assumptions: textual data in qualitative research are considered particularly suitable for revealing specific aspects of the phenomenon under investigation; furthermore, this phenomenon is assumed to exhibit patterns that are traceable within the data themselves and require interpretation.

Thus, reality is assumed to manifest through patterns in textual data. Texts are not merely surface descriptions but traces of deeper cognitive or social structures. This connects to the interpretivist paradigm, in which meaning is accessed through language.

With respect to these types of data, the authors note that:

"As these methods require a series of judgments by the analyst, they are difficult to validate and reproduce. Further, they are hard to scale so that they are unavailable to the analysis of large-scale data, e.g., chats in online learning environments."

The phase of pattern interpretation is laborious and requires numerous judgments. This is likely due to the richness of the contexts in which such phenomena occur, or because pre-existing theories and contextual information must be taken into account to extract meaningful patterns. However, this work faces clear limitations in settings such as digital learning environments. These environments are ontologically particular because reality is dual: it consists of both the physical, "real" world where the student is situated, and the virtual, online space where the student acts and interacts through the platform. The virtual thus becomes part of reality itself. In this context, large-scale textual and numerical traces are produced and regarded as valuable sources of information for accessing and understanding students' behaviour. All these elements are ontologically valid, blending offline dimensions with digital traces.

Furthermore, previous excerpts suggest that the type of phenomena under study sometimes emerges from large datasets, which challenge qualitative methods mainly because of their scale. Therefore, ontologically speaking, within CGT reality is conceived as having a dual nature: (i) formal regularities detectable through computational methods applied to large datasets, and (ii) contextual meanings interpretable only through human understanding. Ontology here bridges post-positivist (regularities) and interpretivist (meanings) perspectives.

Epistemological commitments

The epistemological assumptions of this study revolve around how knowledge is generated through the interplay between human interpretation and computational detection. Since CGT can be intended as an intentional evolution of GT, the authors first explicitly describe the epistemic process of qualitative methodologies such as GT as consisting of two main steps: "recognizing patterns in the data and interpreting these patterns". Thus, in that context, knowledge is assumed to emerge from the dual movement of identifying regularities and interpreting them. The epistemic process is therefore both empirical (rooted in the detection of recurrent structures within data) and interpretive, since meaning is constructed through theoretical and contextual reflection.

In the Computational Grounded Theory (CGT) framework, this epistemic movement is conducted following the similar phases, where the machine mainly support specific ones:

"CGT proceeds in a process of three consecutive steps. […] computational techniques for pattern detection; human interpretation for quality and depth; computational testing for robustness across the dataset."

Knowledge is thus conceived as progressively constructed through iterative human–machine steps that combine detection, interpretation, and validation. Each stage of the process informs the next: the outcomes of computational detection are reinterpreted by the human analyst, whose refinements are then re-tested through computational validation. This cyclical logic defines knowledge production as a dynamic process of negotiation between empirical regularities and interpretive depth, which is different from GT for letting the machine explore data, giving value to computational capabilities to do so on a large scale.

The action of humans and machines are explicitly described as integrative and complementary. In fact, *"CGT does not aim at simply automating parts of the qualitative process by using AI, but rather aims at integrating AI into the human analyst's workflow within a qualitative analysis"*.

This statement conveys an important epistemic position: artificial intelligence is not a substitute for human reasoning, but is conceived as an extension of it. Also, the machine's computational power and capacity to detect latent patterns are intentionally paired with the researcher's theoretical sensitivity and contextual understanding.

This complementarity is not incidental but constitutive of the epistemic design. The human and computational components carry distinct yet interdependent knowledge-making capacities—pattern recognition on the one hand, and contextual interpretation on the other. Their collaboration produces an expanded epistemic system that "*is quantitatively and qualitatively different from what a human or machine can do alone*". However, it is important to underline that, since computational models operate through statistical rules valid only across large datasets, the patterns they produce correspond to regularities observable at scale. These regularities constitute the empirical foundation of the analysis: unlike in traditional Grounded Theory, where initial patterns may emerge from small samples without statistical consideration, in CGT the starting point is a set of statistically supported structures. Yet the novelty of CGT does not lie solely in this quantitative robustness, but in the hybridity of the epistemic operation itself. The resulting insights are not merely faster or larger-scale versions of those generated by qualitative analysis; they are qualitatively distinct, emerging from the dialogical engagement between algorithmic modelling and interpretive reasoning, where statistically grounded regularities are reinterpreted through human theoretical sensitivity.

Another central feature of this knowledge production is iterativity. The CGT process is described as inherently recursive:

"These first two steps are potentially iterative […] incrementally refining the interpretations of patterns until convergence is reached."

Knowledge is therefore not produced in a linear sequence but through continuous refinement and exchange. Meaning emerges through an evolving dialogue between machine-generated outputs and human interpretation, where each informs and constrains the other. Convergence is reached not by replacing human judgment with algorithmic certainty, but by aligning both through iterative cycles of sense-making.

Finally, the study explicitly addresses the question of validity and confirmation:

"The last step, pattern confirmation, aims at testing that the patterns found and interpreted […] are not an artifact of a specific NLP or ML technique or a biased interpretation by the human analyst".

Here, the epistemic commitment lies in the requirement for methodological self-validation. Knowledge is considered robust only when it can withstand verification across different analytical layers: computational and interpretive. The confirmatory phase thus plays an epistemological role analogous to triangulation in qualitative research. It ensures that findings are not artifacts of the analytical process itself, but stable insights grounded in both empirical consistency and theoretical coherence.

Taken together, these elements define the epistemology of CGT as procedural, integrative, and reflexive. Knowledge is conceived as emerging through the complementarity of human and machine reasoning, iteratively refined and empirically validated. A genuinely hybrid epistemic process that redefines how data-driven theory can be constructed in SER.

Axiological commitments

The axiological dimension of the study concerns the values that guide the methodological design of Computational Grounded Theory (CGT) and the broader implications these values carry for research practice in science education.

A central value of CGT lies in the deliberate and reflective integration of automation within qualitative inquiry. This integration becomes particularly valuable when working with large datasets, where manual analysis would be excessively time-consuming and resource-intensive. At the same time, the study conveys a broader sense of responsibility: efficiency and scalability must be balanced with an awareness of energy consumption and ecological impact. The value of *efficiency*, expressed in faster and more extensive analyses, cannot be detached from the ethical imperative to minimise resource expenditure and environmental footprint. As the authors acknowledge:

> *"A potential caveat here is processing time… computation on very large datasets or using complex ML techniques can require a lot of time or money for powerful hardware, which also taxes the environment".*

However, these values appear to be assumed as secondary and consequential, rather than as the primary values that guide the methodology. Indeed, *"CGT does not aim at simply automating parts of the qualitative process by using AI, but rather aims at integrating AI into the human analyst's workflow within a qualitative analysis".*

Automation is not pursued as an end in itself, but as a means of enhancing the analytical process. The value is thus *integration*, not *replacement*. The machine's efficiency and consistency are mobilized to extend human analytical reach, while the researcher's interpretive awareness remains essential. This expresses an ethic of *collaboration between human and machine*: a balanced distribution of epistemic authority that reflects a commitment to methodological pluralism and cognitive complementarity.

Behind this automation lies a methodological concern that is particularly crucial in science education research: the ambition to make qualitative analysis more scalable, reproducible, and methodologically rigorous. The authors highlight that:

> *"In this way, CGT aims at addressing questions about validity, reproducibility, and scalability in qualitative research while preserving the theoretical sensitivity and unique inferencing capabilities of the human analyst".*

Here, values traditionally associated with quantitative paradigms—*validity*, *reproducibility*, and *rigor*—are recontextualized within a qualitative framework. Computational support enables researchers to work with large-scale textual datasets, overcoming the limitations of resource-intensive manual coding. In fact:

> *"Computational guidance in CGT allows us to engage with very large data where lacking resources prohibit or limit traditional qualitative approaches".*

CGT is explicitly developed to introduce mechanisms for *transparency*, avoiding researcher bias and human error, which are often considered weaknesses of qualitative analysis. This occurs through the transformation of human analytical actions into algorithmic procedures that can be documented and repeated:

> *"The usage of computational tools in the steps of CGT partly address this issue [reproducibility]. There are human decisions and actions involved… which must be made transparent by the analysts for the sake of reproducibility".*

Thus, the axiology of CGT embodies a dual aspiration: to retain the interpretive depth of qualitative research while attaining the reproducibility and scalability characteristic of data-intensive approaches. In other words, a

value system that reconciles the often-opposed ideals of qualitative and computational research (pattern recognition abilities of computers and theoretical sensitivity of the human analyst) is promoted.

This reflects a *hybrid axiology*, where the human and the machine are regarded as complementary contributors to the production of rigorous and meaningful knowledge. Efficiency, reproducibility, and formalisation coexist with interpretive sensitivity, reflexivity, and contextual awareness. In this sense, CGT embodies a new methodological ethic: not the subordination of one value system to another, but the *coexistence* of diverse methodological virtues within a shared analytical process.

Taken together, these axiological commitments reveal the value framework of CGT as integrative, reflexive, and responsible. The method embodies a balanced ethics of automation, seeks transparency and reproducibility without sacrificing theoretical depth, embraces plural methodological values, and remains attentive to the material and environmental implications of computational research.

5.5.3 Case 3: Network Thematic Analysis

The third case is the article *"Identification of positions in literature using thematic network analysis: the case of early childhood inquiry-based science education"* by Mariegaard, Seidelin, and Bruun (2022). This study introduces *Thematic Network Analysis (TNA)*, a hybrid methodology that combines thematic analysis with network analysis.

The method proceeds through several steps:

1. **Corpus selection**: 35 peer-reviewed articles on inquiry-based science education in early childhood were identified through systematic criteria.
2. **Extraction of theoretical passages**: researchers selected excerpts that explicitly articulated theoretical positions on inquiry.
   Construction of linguistic networks: the excerpts were preprocessed, grammatically reduced, and converted into weighted lexical networks where nodes represent words and links represent co-occurrence.
3. **Clustering and mapping**: community detection algorithms were applied to identify clusters of words, interpreted as candidate themes.
4. **Qualitative revision**: thematic maps were compared back to the original texts and critically refined through iterative naming and interpretation.

The case focuses on early childhood inquiry-based science education but opens broader methodological discussions. Tools included Gephi for network visualization and R packages for textual and network analysis, combined with qualitative interpretation of the maps.

This study represents a clear example of mixed methods by enriching traditional thematic analysis (qualitative) with network analysis (that computes quantitative measurements such as distances between entities, showing relational features). It improves transparency and reproducibility while retaining interpretive depth, marking a significant advancement over thematic analysis alone.

Ontological assumptions

In the TNA study, reality in educational research is assumed to be constituted by the theoretical positions that underpin empirical work. As the authors state:

"In educational research, such information includes different positions within a field; for example, theoretical positions underpinning empirical studies".

and further clarify:

"Such theoretical positions may influence interpretations of research results and remain a challenge to map and present".

These excerpts suggest that theoretical positions are not abstract or detached entities but are embodied in texts. Within the context of science education, they exert influence on how research findings are interpreted, often implicitly. They are thus conceived as ontologically real objects, yet complex and multifaceted, frequently remaining hidden and requiring methodological tools to make them visible.

The article also emphasises that educational research is embedded in historical and philosophical traditions, such as Dewey's philosophy in early childhood inquiry-based science education (IBSE). These traditions act as anchoring frameworks for inquiry, but they are neither singular nor stable. Instead, multiple standpoints coexist, sometimes competing with one another, reflecting a plural and contested object of study that demands recognition.

At the same time, the study explicitly embraces a constructivist ontology of language, as illustrated by the authors' citation of Willig (2013):

"However, following Willig (2013), we argue that no reading – and indeed any interpretation – can be said to be 'true' or 'right', since language is constructive and functional".

From this perspective, interpretations are not valued for their truth or correctness, since language itself is understood as a constructive and functional medium. Reality is therefore not independent of discourse but emerges through linguistic practices and interpretive acts. In TNA, the very act of mapping theoretical positions becomes a way of rendering visible how discourse constructs the field of science education, transforming textual traces into representations of the epistemic structures that shape research itself.

Epistemological commitments

The epistemological assumptions of the Thematic Network Analysis (TNA) study revolve around how knowledge is constructed, represented, and validated through the integration of qualitative interpretation and computational modelling.

This methodology uses *"thematic analysis and network analysis to construct maps, analyse and synthesize theoretical positions within educational research"*. The decision to use these two methods together is grounded in the idea that knowledge in educational research does not reside solely in the content of individual texts, but also in the *relationships* among them. In fact, the authors argue that traditional reviews (often based on thematic categorisation) *"may overlook important relationships in the textual presentations, which in turn could have led to different interpretations and theory."* Thus, traditional reviews tend to emphasise thematic content, while overlooking the relational dimension. By contrast, TNA positions relationships between words, concepts, and theoretical statements as epistemically significant. Mapping these relationships enables researchers to access the structure of theoretical positions in a given field, making visible connections that would otherwise remain implicit.

TNA assumes that the complementarity between the quantitative logic of network analysis and the interpretive depth of thematic analysis enables themes to be analysed in a systemic rather than merely categorical way, leading to a more holistic mapping of knowledge. As the authors explain:

"In this paper, we illustrate a novel method of literature review, which combines thematic analysis with recent mapping review techniques by thematic network analysis (TNA)… Our purpose is not only to map and describe themes, but to synthesize, analyse and map different positions in a sub-field of education".

The method relies on the assumption that hidden structures within textual data can become visible through computational modelling, which detects statistical co-occurrences and relational patterns. These quantitative representations do not replace qualitative interpretation but provide a complementary lens that enhances transparency and reproducibility. In fact, "*researchers using the same data set and the same method should arrive at the same results (i.e. thematic networks)".*

TNA also embodies a model of *integration by iteration*. The process of analysis is explicitly described as non-linear, involving a continuous dialogue between human and computational operations:

"TDNA (Thematic Discourse Network Analysis) is an iterative methods approach to analysing and visualizing relationships in textual data… our movement from phase to phase was not [linear]".

The authors describe decisions such as grammatical reduction, parameter settings, or the refinement of clusters are informed by evolving qualitative insights. Conversely, network outputs guide further interpretation and revision. This iterative exchange exemplifies an epistemic process in which human theoretical sensitivity and computational detection mutually inform each other, gradually converging toward stable and interpretable representations. As the authors put it:

"Our proposed method can be seen as an integration of qualitative (thematic analysis) and quantitative (network analysis) methods. We see the two as being on equal footing in the method, because they iteratively inform each other throughout the process".

The authors also make clear that, despite its computational components, the TNA approach maintains a strong recognition of the human researcher's central epistemic role. Human judgment is indispensable in refining categories, adjusting analytical rules, and interpreting emergent structures. Triangulation, both among researchers and across computational outputs, is used as an epistemic safeguard to ensure reliability and to verify whether network representations correspond to the theoretical positions recognised by the field. As they note: *"A further increase of reliability might be to have researchers work independently on different aspects of the method… If researchers' analyses converge over time, then one could say that the results were reliable".*

Finally, TNA is underpinned by a reflexive epistemology that acknowledges how methodological choices directly shape analytical outcomes. Every decision, from data preprocessing to community detection, affects the resulting networks and, consequently, the narrative constructed from them. For this reason, *traceability* is an essential requirement: researchers must document how analytical rules are set and how changing them alters the resulting maps. As the article states: *"Therefore, keeping track of rules and how changing them changes network representations are essential to interpretations and the final narrative".*

Reflexivity and transparency therefore operate as twin epistemic conditions that sustain the validity of the method and enable its reproducibility.

Taken together, these commitments portray TNA as a deeply integrative epistemic framework. Knowledge is conceived as relational, mediated by the complementary operations of computation and interpretation, and achieved through iterative refinement and reflexive awareness. Rather than subordinating one tradition to the other, the epistemology of TNA exemplifies a balanced and dialogic form of mixed inquiry, where quantitative precision and qualitative insight converge to produce a richer and more accountable understanding of educational knowledge.

Axiological commitments

From the analysis of this article, Thematic Network Analysis (TNA) study seems resting on three central values (: transparency and reproducibility, the responsibility and interpretive centrality of the researcher, and the balance between computational modelling and contextual meaning.)

First there is a core value underlying TNA is the commitment to transparency and reproducibility, both understood as ethical and methodological imperatives. The authors recognise that while thematic analysis offers great flexibility, it is *"hard, if not impossible, to reproduce, and great care must be taken to ensure transparency"*. TNA responds to this challenge by introducing computational formalisation and by openly sharing its materials. In fact:

> *"R code, data and instructions for reproducing our results are available"*.

These practices express a collective ethos of accountability: knowledge gains credibility also from the possibility of being re-examined, replicated, and refined through shared procedures.

Secondly, despite the presence of algorithmic processes, the study reaffirms the central role of the researcher. Human judgment is indispensable in defining parameters, adjusting analytical rules, and interpreting outcomes. The authors note that *"the design and implementation of algorithmic rules are at the heart of the researcher's decision process. These rules shape thematic networks, thereby framing the whole interpretative process"*. This awareness reflects an ethic of responsibility: computational tools expand analytical reach, but methodological choices as well as interpretation remain a human act that requires reflexivity and theoretical sensitivity. The researcher's value lies not in delegating meaning-making to algorithms but in orchestrating a transparent and thoughtful interaction between human reasoning and machine computation.

Finally, the study promotes the value of *balance*, between the formal power of the model and the contextual depth of interpretation. The computational model allows for systematic and extensive measurement of relationships, offering a structured map of distances and connections across textual data. Yet the researcher determines the appropriate granularity that lets meaningful themes and relational links emerge. This is illustrated when the authors write that "the purpose of this phase was to reduce the complexity of the network representation of the data while still preserving critical relationships" and also *"In reporting our findings, we found that one theme was too prevalent […] Therefore, in later analyses, we removed this theme from the thematic map and worked with the most prevalent remaining connections"*. Such decisions reveal an ethical concern for proportionality: the simplification required by modelling should never come at the expense of interpretive richness.

Taken together, these axiological commitments portray TNA as a methodology grounded in openness, responsibility, and equilibrium. It strives for reproducibility without rigidity, embraces computational precision without displacing human insight, and seeks a dynamic balance where methodological transparency and interpretive meaning mutually reinforce each other.

5.6 Insights from expert interviews

5.6.1 Machine Learning and NLP with Dr. Peter Wulff

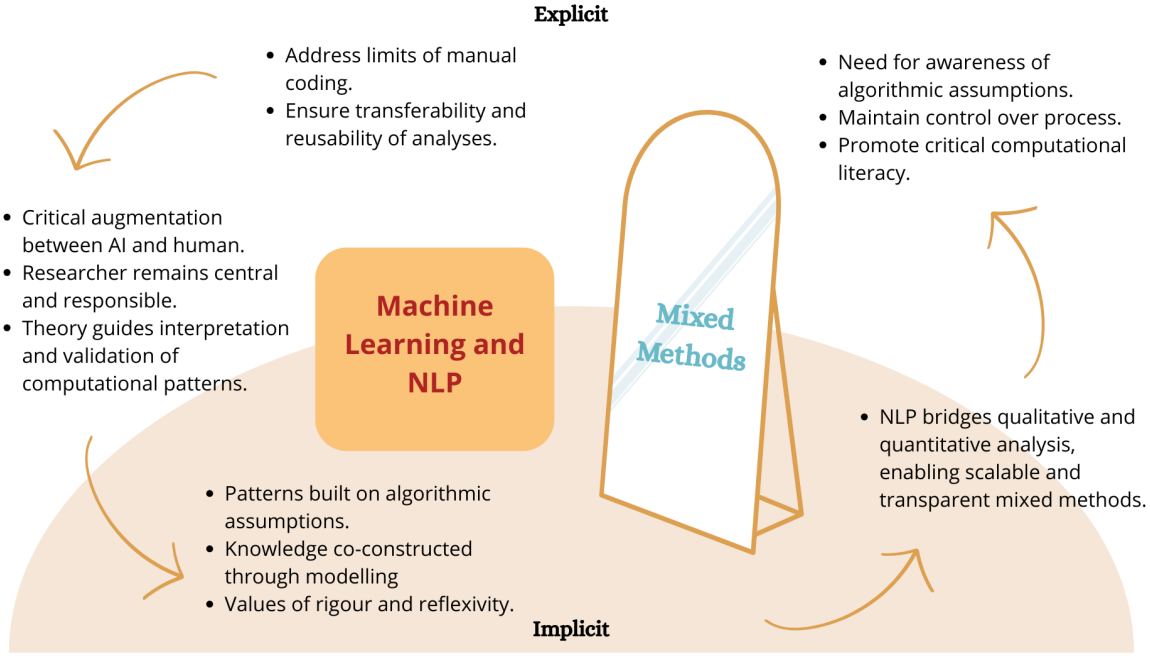

Figure 5.3 - Insights from the interview with Dr. Peter Wulff about the use of ML and NLP in Science Education Research

(a) Researcher background and motivations

Dr Peter Wulff's academic trajectory reflects a continuous attempt to bridge disciplinary, methodological, and representational boundaries. Trained in physics and German linguistics, he pursued a PhD on gender issues in physics, investigating students' identity development over time through longitudinal quantitative analyses. After the doctorate, his work expanded toward teacher knowledge in physics education, where he began collaborating closely with computer scientists and computational linguists. It was in this context that he first adopted machine learning and natural language processing (NLP) techniques for the automated classification of teachers' written reflections—an experience that revealed both the potential and the limits of computational text analysis.

Since 2014, Wulff has been deeply involved in science education research, focusing increasingly on the methodological implications of analysing text as data. Although his background is predominantly quantitative, he recognised the interpretive complexity of language-based research, particularly in studies of teacher reasoning and problem solving. As he observed, textual data are "so interconnected and multifaceted" that conventional coding often struggles to capture their full meaning.

This awareness led him to view NLP and ML as bridging tools between quantitative and qualitative reasoning. He described them as fundamentally quantitative methods that, through topic modelling and clustering, "can enrich quantitative data to be more of a qualitative sort." The motivation to adopt these approaches arose from a double need: to address the practical limits of manual coding—its dependence on human coders, its fragility over time, and its limited reproducibility. He later reflected that this limited reproducibility is also influenced by cultural factors: few researchers seem to be willing or engaged to re-use coding rubrics developed by others, which further hinders the scalability and adoption of qualitative research practices. Another motivation to adopt computational methods is to preserve the analytical knowledge produced through coding for future reuse. As he noted, once a team of coders or assistants leaves a project, "it is very difficult for another human being to reconstruct the process of how the data was classified and coded." Computational models, by contrast, offer the possibility of transferability—making the logic of classification explicit and reusable across contexts.

While acknowledging that "the computer is not a good and deep qualitative researcher yet," Wulff envisions AI-based methods as a critical support system for human inquiry, expanding the range of data that can be explored and improving methodological transparency. In this sense, his motivation to adopt ML and NLP was not merely technical but epistemic: a search for methods that can retain the richness of qualitative interpretation while ensuring the reproducibility and scalability that educational research increasingly demands.

(b) Roles of the human researcher, the computational model, and theory

In describing the interaction between human reasoning and computational modelling, Wulff characterised them as *"totally different means of analysis."* Qualitative inquiry, he explained, is inductive and sensitive to ambiguity, while machine learning provides systematic means to detect latent regularities across large and complex datasets. The two forms of analysis are thus not substitutes but complements, each addressing the other's limitations. Human researchers struggle to identify patterns within vast textual corpora, whereas computers can efficiently process them—albeit only within the assumptions and parameters defined by their algorithms.

This operational asymmetry defines the logic of collaboration: the computer *detects*, while the human *interprets and evaluates*. For Wulff, this interplay demands a continual process of *triangulation*, in which researchers examine the assumptions guiding algorithmic models and assess how these shape the patterns they return. The aim is not to delegate interpretation to machines but to extend human capacity for analysis under explicit methodological control.

He described this relationship as one of critical augmentation—a partnership in which artificial intelligence enhances human analytical power while leaving the researcher at the centre of decision-making. The human, he argued, must remain the *responsible agent* of the process, both for ethical and legal reasons, as emerging frameworks such as the EU AI Act explicitly require human oversight in research and education. AI should therefore be seen as a methodological companion: a system that amplifies human perception and supports analytical transparency without displacing human judgement.

While emphasising this synergy, Wulff also noted the potential for distraction and distortion, particularly when using large language models whose generative behaviour can be unpredictable.

Such awareness underscores his broader stance that methodological innovation must be accompanied by critical vigilance—an attitude of engagement that balances curiosity with responsibility.

(c) Epistemological, ontological, and axiological assumptions

At the epistemic level, Wulff emphasised that computational models do not discover meaning but detect *patterns of association* determined by the assumptions embedded in their algorithms. These groupings are not neutral reflections of reality but the outcome of pre-specified rules. Understanding their significance therefore requires human triangulation and interpretation—a process that transforms statistical regularities into meaningful insights.

From this perspective, knowledge is produced through a dynamic exchange between detection and interpretation, automation and reflection. Wulff described this as a *construction cycle* in which patterns both exist in the data and are shaped through analytical practice. *"From an ontological perspective, the patterns exist in the data,"* he explained, *"but they are also constructed, in the sense that humans provide the algorithms to detect them."* The act of modelling thus transforms potential structures into observable forms—*"we enabled the machine to give us back these patterns."* In this view, knowledge emerges from iterative translation between representational systems: human, linguistic, and computational.

This epistemic stance positions AI neither as an objective observer nor as a mere extension of human subjectivity, but as a *mediating system* that materialises the interpretive assumptions embedded in its design. By making these assumptions explicit, AI-based methods render the process of interpretation more transparent. Wulff regarded this as an epistemic advantage: unlike manual coding, where biases remain implicit, computational modelling encapsulates analytical decisions within a reproducible framework. Yet transparency, he cautioned, does not equate to neutrality—algorithms encode their own forms of bias and can perpetuate or amplify them if not critically examined.

The role of theory further reinforces this constructivist orientation. Wulff argued that AI can actively contribute to theory-building in empirical sciences by revealing relationships that humans might overlook, but that theoretical frameworks must still guide interpretation. Theories act as *filters of relevance*, defining which patterns are meaningful and ensuring that computational findings are grounded in disciplinary understanding.

Axiologically, Wulff's position is underpinned by values of rigour, reflexivity, and responsibility. Rigour arises from reproducibility and explicitness; reflexivity from the continual awareness of the assumptions guiding both human and algorithmic reasoning; and responsibility from the ethical requirement to maintain human agency at the centre of interpretation.

(d) Positioning ML and NLP in relation to mixed-methods research

Wulff's conception of mixed methods departs from the traditional view of combining two distinct methodological strands and instead focuses on the transformation of qualitative data into a form that can be quantitatively interrogated without losing its interpretive grounding. While he acknowledged the classical typologies of mixed-methods designs—explanatory, exploratory, and embedded—he emphasised that the deeper challenge lies in creating a genuine bridge between the qualitative and the quantitative.

In this regard, he viewed natural language processing as a methodological interface that makes such integration possible. By enabling the quantitative representation of language data, NLP allows textual analysis to be *triangulated* with other forms of quantitative evidence, thereby expanding the analytical possibilities of educational research. As he explained, conventional qualitative content analysis already performs a rudimentary form of mixing when it categorises textual material and correlates those categories with numerical variables. However, he regarded this process as *"cumbersome, unscalable, and inconsistent,"* often depending on interpersonal negotiation rather than methodological rigour.

Through computational modelling, by contrast, this integration becomes both *scalable* and *explicit*. Text can be systematically transformed into structured numerical representations that retain semantic information and can be correlated with other quantitative measures. The result, Wulff argued, is not a replacement of qualitative depth but a reconfiguration of how it connects to empirical measurement. The process remains interpretive—humans still decide which features to analyse and how to read the resulting patterns—but it occurs within a framework that allows reproducibility and cumulative knowledge building.

At the same time, Wulff maintained a cautious stance regarding the limits of this methodological convergence. The statistical representations produced by NLP may lack the depth and contextual sensitivity achieved by traditional qualitative analysis. For him, the epistemic value of these methods lies not in dissolving the distinction between qualitative and quantitative, but in *making their interaction more rigorous and transparent*. In this sense, NLP-based approaches exemplify a new generation of mixed methods: those in which integration is not merely additive but structurally embedded in the analytical process itself.

(e) Applications, competences, and potentials for Science Education Research

In discussing the future of machine learning and natural language processing within science education research, Wulff emphasised that the essential competence for researchers is not technical mastery but *methodological awareness*. Rather than learning to build algorithms from scratch, scholars should cultivate a critical understanding of how these systems work—the assumptions that guide them, the types of data for which they are suited, and the conditions under which they yield reliable results. As he put it, understanding ML and NLP means knowing *"what are the assumptions that go into it, what kind of phenomena you can analyse with them, and under what circumstances they have worked in the past."*

This reflective understanding forms the basis of methodological control. Wulff warned that outsourcing analyses to external services risks detaching researchers from the epistemic core of their work. Instead, he encouraged direct engagement with open-source or locally implemented models, where scholars can explore, test, and adjust parameters: *"It's very important that you have a little bit of control over the process… the analysis of the data is so important and so crucial for your conclusions."* Such control ensures transparency and accountability, both of which are central to the integrity of AI-supported inquiry.

To illustrate why this awareness matters, Wulff offered concrete examples of algorithmic assumptions. In the common *bag-of-words* model, for instance, word order is ignored—making the statements *"current creates voltage"* and *"voltage creates current"* indistinguishable, despite their different conceptual meanings in physics. This simple case demonstrates how seemingly technical choices can have profound implications for the kind of knowledge that can be produced. Similarly, he warned that

large language models are trained on vast public datasets containing misconceptions and inaccuracies. Researchers using such systems in educational contexts must therefore design safeguards to prevent these errors from reappearing in automated feedback or tutoring applications.

For Wulff, these considerations outline a new form of *computational literacy* for science education researchers—one grounded not in coding proficiency but in epistemic sensitivity and critical oversight. The value of AI lies in its potential to make analysis scalable and transparent, but only if researchers remain aware of how meaning is mediated by the assumptions, architectures, and training data of the models they employ. Such methodological reflexivity, he concluded, is indispensable for ensuring that future uses of AI in education remain both scientifically rigorous and ethically responsible.

5.6.2 Computational Grounded Theory with Dr. Marcus Kubsch

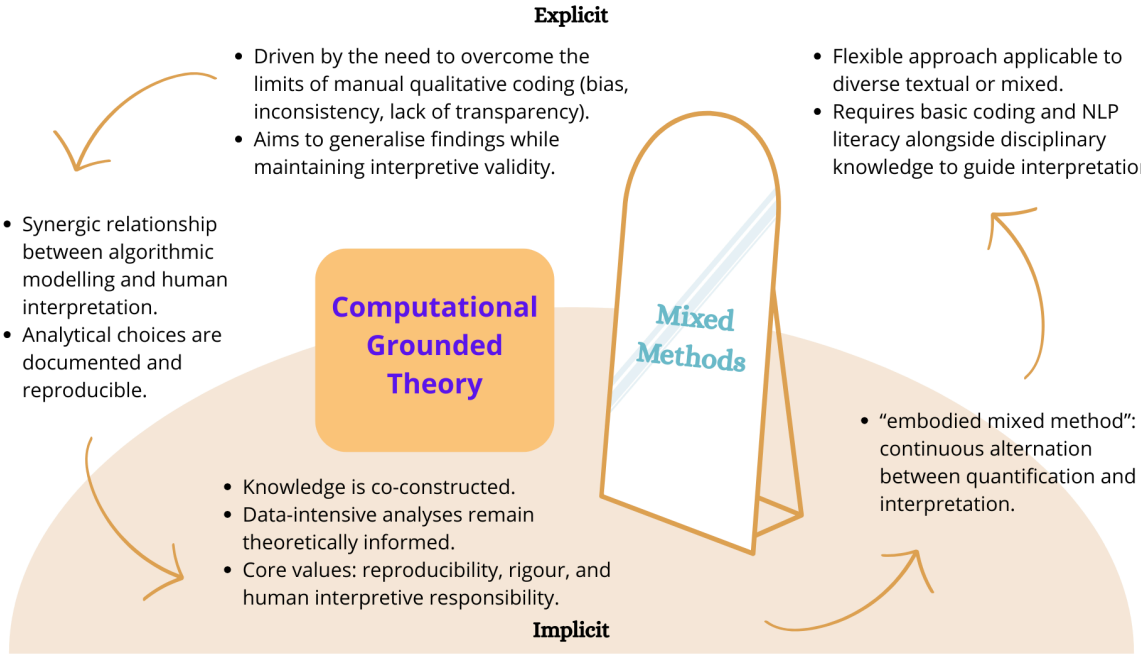

Figure 5.4 - Insights from the interview with Dr. Marcus Kubsch about Computational Grounded Theory in Science Education Research

(a) Researcher background and motivations

Dr Marcus Kubsch described a trajectory that bridges the empirical precision of physics with the interpretive demands of education research. After initial work in astrophysics he "*switched over to physics education research"*, pursuing a PhD that *"combined these aspects of physics, but also issues around language"*. This transition was facilitated through an international collaboration with Michigan State University, where he began to investigate how students learn and use scientific language.

In his doctoral and postdoctoral work, Kubsch's research focused on *open-ended questions* designed to capture how learners reason about energy concepts. *"We worked a lot with open-ended kind of questions because we wanted to assess this kind of deep learning in the sense of the NGSS",* he

explained. *"You cannot really assess this with a multiple-choice question, but then the issue is that you get a lot of text data"*. Managing and analysing such unstructured responses introduced him to the methodological challenges that would later motivate his turn to computational approaches.

As he recounted, *"I was involved in designing the coding manual, supervising these processes… and that got me interested in doing this more reliably"*. The experience of coordinating qualitative coding illuminated how inter-coder reliability and implicit interpretive choices could compromise transparency. *"You always have this issue with multiple people coding the same data… is it reliable, yes or no? Can you somehow get to a more objective standard?"* These questions became the starting point of his methodological reflection. In 2016-2017, he began *"playing around with topic modelling and similar approaches,"* experimenting with computational tools that could support more systematic forms of textual analysis.

Over time, this methodological curiosity developed alongside his three major research interests: learning progression analytics, trust in science and uncertainty, and a persistent focus on the methodological foundations needed to address both. *"For me,"* he observed, *"thinking about research methods is always driven by the kind of problems that I face, and in that sense, I very much care about using methods appropriately"*. His physics background instilled a set of epistemic values that continued to shape his work: *"If you don't have good measurements then whatever comes out of that doesn't really matter… I've seen in the literature a lot of issues of reproducibility and replicability… and if we don't do this well, then all the other stuff doesn't really matter and isn't really science."*

The adoption of Computational Grounded Theory arose directly from these concerns. Traditional qualitative approaches, he admitted, often left him *"a little uncomfortable"* because of their dependence on tacit interpretive choices: *"Is what I'm finding really a substantive pattern? Can it generalise or not?"* Coding manuals, he noted, are difficult to reproduce across researchers and contexts because *"many of these choices are never recorded"*. For him, CGT offered a methodological remedy: a way to *"document decisions by the kind of analysis script you're using"* and to make reasoning explicit and reproducible. *"They may not be more objective necessarily, but they are reproducible… if I use the same method on the same data, there will be the same outcome."* In his words, this transparency is not only technical but ethical. His commitment to CGT reflects not only a technical preference but a deeper epistemic stance, an attempt to align the practices of qualitative inquiry with the methodological ideals of scientific reliability.

(b) Roles of the human researcher, the computational model, and theory

In describing how the human and computational elements interact in CGT, Kubsch emphasised both their distinct capacities and their mutual dependence. The computer, he explained, offers a kind of procedural stability and endurance that humans cannot achieve: *"Another benefit is that we have all these issues, especially if you have large data sets, with drift effects in coding. You start coding something and then after some time there are subtle changes—people become more familiar, or they code differently in the beginning than in the end. The computer, by contrast, outputs in the same way all the time."* This mechanical consistency also provides a safeguard against idiosyncratic interpretation: *"If you cannot replicate a finding, we can at least be sure that this is not due to another grad student coding the data in some idiosyncratic way, because the computer is doing the same thing all the time."*

Beyond this stability, the computer extends the scope of human perception by overcoming the cognitive limits of attention and memory. As Kubsch observed, *"Our attention is limited… there is*

only a certain amount that I can store in my working memory to identify patterns across hundreds and hundreds of texts. That is very different for the computer, and in this way low-frequency phenomena have a higher chance of being detected." Subtle regularities that a human analyst might overlook thus become visible through computational assistance.

At the same time, Kubsch resisted the idea of automation as replacement. The most productive analyses, he argued, arise from a *"synergetic combination of using computational approaches and human intelligence."* The computer can detect patterns, but only the human researcher can evaluate their meaning: *"The computer may pick up patterns that a human never would, but then you as a human can bring your substantive knowledge and theoretical background to this to make meaning of it."* This complementary relationship ensures methodological balance: *"You avoid too much subjectivity because the computer can verify against this and see—Is this pattern really there?—and on the other hand, you avoid having the computer just picking up stuff that may be there but is not meaningful for what we're interested in, because you have the human in the loop."*

Kubsch described the interaction between human and computational agents as partly sequential and partly dialogical. Certain stages, such as model design and data preparation, require the researcher to make key decisions in advance: *"In the beginning, as the researcher, you decide what data to use, what computational approach to employ, and there are a lot of ways to represent text and a huge class of modelling approaches. This is where it's pretty much a sequence—you're making these decisions."* However, once the models generate candidate patterns, the process becomes dialogical: *"In this pattern refinement step the computer puts out some patterns and then you are getting into reading these texts that the computer identified as having something in common. This is where it becomes more dialogical… you read and may say, is this really different? How far away are these things? If I play around with the hyperparameters, do two things suddenly become one or not?"* This *back-and-forth* cycle of exploration and adjustment is, for Kubsch, the heart of CGT: a process in which meaning and structure co-evolve through continuous negotiation between algorithmic detection and human interpretation.

Finally, Kubsch reflected on the inescapable role of theory within this process. He rejected the notion of a purely data-driven analysis, noting that *"you cannot look at data without pre-existing ideas… you will always bring something to this, advertently or inadvertently."* For him, CGT reveals rather than eliminates this interpretive presence: *"Theory can never play no role—there's never really theory-free, just data-driven. The data never just speaks for itself."* In practice, computational analysis can even challenge theoretical assumptions by producing patterns that do not align with established frameworks: *"If it tells me these texts have something in common and it doesn't fit my theory, I have to find a convincing argument—either why this pattern doesn't matter, or I may need to revise my theory and add something else. And this doesn't happen if you just do this with a human coder; there is nothing that forces you to argue with."* In this sense, the algorithm functions as a *counter-voice* within the analytical dialogue, inviting theoretical revision and reducing the risk of confirmation bias.

(c) Epistemological, ontological, and axiological assumptions

Kubsch's reflections on knowledge and method reveal a balanced epistemic position in which interpretive openness and methodological rigour coexist. At the core of this stance lies an understanding of reality as *co-constructed* through the interaction between human interpretation, language, and analytical procedure. As he explained, *"I think it's a co-construction really… interpretation of text is not objective in that way, as there is no true meaning inscribed through a*

text." Drawing on the linguistic tradition of Saussure, he added, *"There is no fixed relationship between the signifier and the signified… what we make of it is based on the interpretation of us as linguistic beings in a specific culture and context at a specific time."* From this view, knowledge is not a mirror of reality but a situated representation: *"People using the same approach ten years later may come to different conclusions because they interpret the same pattern differently."*

This constructivist understanding of meaning does not, however, entail methodological relativism. Kubsch emphasised that computational approaches introduce a valuable form of procedural stability into interpretive research. *"There's this certain element of reproducibility,"* he noted, *"because all the analysis is encapsulated by the analysis code that you're using, and so you have this built-in mechanism that is in a way always checking against becoming too associative in your interpretations."* The algorithmic structure functions as an anchor for reflexivity: while interpretation remains contingent, the analytical pathway is traceable and repeatable. Thus, reproducibility becomes not a claim of objectivity, but a practice of *accountable transparency* that safeguards interpretive reasoning from arbitrariness.

For Kubsch, this interplay between interpretive flexibility and computational rigour is what gives CGT its epistemic distinctiveness. *"The benefit of using something like Computational Grounded Theory,"* he observed, *"is that you're getting some of the rigour aspects that this strict methodology… provides, while maintaining all that makes qualitative research desirable in the first place—you can bring these different perspectives to bear and use your theoretical interpretation and interpretation abilities here."* CGT, in his account, bridges two epistemic cultures: it maintains the *plurality and contextual sensitivity* of qualitative inquiry while embodying the *replicability and procedural discipline* of quantitative research.

Axiologically, Kubsch's position is grounded in values of *rigour, reproducibility,* and *human interpretive responsibility*. The computational component provides what he called a "built-in mechanism" for testing the coherence of interpretation, yet the meaning-making act remains a human and cultural one. Knowledge is valuable not because it is fixed or absolute, but because it is *documented, reflexive,* and *open to revision*. CGT thus enacts a methodological ethic: to pursue reliable and transparent inquiry without effacing the interpretive and contextual dimensions of understanding.

(d) Positioning CGT in relation to mixed-methods research

When asked whether Computational Grounded Theory could be considered a mixed-methods approach, Kubsch offered a nuanced interpretation that reframes what "mixing" means. He distinguished between multimodal analyses, where different data sources are simply combined, and genuinely integrative approaches that merge distinct ways of reasoning. From that perspective, CGT could indeed participate in such designs, but its most distinctive character lies elsewhere.

For Kubsch, CGT embodies a deeper integration of epistemic logics. *"At the beginning of each and every quantitative analysis,"* he explained, *"there's this implicit answer to the question: is what I'm looking at actually quantifiable? And you're implicitly saying yes if you pursue this"*. CGT operates in this threshold space, *"because in a way we need to quantify the text into something to make this work. But then we're also bringing the human part to this."* From this perspective, *"you could say this is the very embodiment of using mixed methods, because we're quantifying something, but we're not stopping there"*. The integration occurs not through the juxtaposition of separate procedures but through a continuous alternation between quantification and interpretation.

Kubsch further reflected on the evolving meaning of quantification, noting that, much like in physics, measurement in language modelling extends beyond assigning single numerical values. Representing words through high-dimensional vectors transforms quantification into a complex, multidimensional description—one that increasingly blurs the traditional boundary between qualitative interpretation and quantitative representation. As he observed, with the rise of advanced natural language processing *"maybe the distinction between qualitative and quantitative starts to blur."* This view encapsulates what might be called the embodied epistemology of CGT: a method where meaning and measurement intertwine within the same analytical act.

Yet this synthesis also demands vigilance. Kubsch warned of the dangers inherent in both traditions: *"There is danger in both traditional qualitative and quantitative research—you can get carried away by it."* Numbers can convey an illusion of precision: *"If we quantify things, it always seems scientific, seems rigorous… but someone looking at this qualitatively might say your test measured something that doesn't matter."* Conversely, overemphasising the individual case risks anecdotalism: *"This can all be very true, but it does not explain overall how teaching and learning of science works."* For him, CGT's value lies in balancing these extremes, building models that *explain and predict how learning occurs* while *remaining sensitive to particularities*.

Finally, Kubsch recognised a further risk that comes with the scaling capacity of computational methods: *"If we start to scale what we're doing… the idiosyncrasies can be washed out by the majority."* The loss of nuance, he cautioned, is a genuine epistemic cost of large-scale modelling. *"These small phenomena can get blurred in the patterns of the majority,"* he warned, and even the use of large pre-trained language models introduces opacity: *"We don't know what goes into them or what is represented in them."* Such reflections reveal that, for Kubsch, embracing the computational turn also entails maintaining critical awareness of its limitations.

Through this perspective, CGT emerges as a *reflexive, embodied form of mixed methods*—a methodological space where the formal rigour of quantification and the interpretive depth of qualitative reasoning co-exist within the same epistemic act. The distinction between the two is neither erased nor hierarchically ordered but enacted as a living dialogue, one that must remain transparent, critical, and open to revision.

(e) Applications, competences, and potentials for Science Education Research

In his reflections on the application of Computational Grounded Theory within science education research, Kubsch highlighted both the methodological demands of the approach and its versatility as a broader mode of inquiry. He described CGT not merely as a tool but as a *way of approaching problems* in a data-informed and reflexive manner. *"You can use this kind of approach to really nail down in a data-informed way a rubric,"* he explained. *"The outcome of Computational Grounded Theory doesn't always have to be the answer to a substantive research question, but it can also help you to develop something that you need as a part of your process."* In this sense, CGT extends beyond a specific analytic pipeline: it represents a methodological stance that can support both theoretical exploration and practical development within research and teaching.

Implementing CGT demands a combination of technical and conceptual competences. Researchers should possess a basic understanding of coding and natural language processing, sufficient to manage and adapt existing analytical scripts, together with a strong grasp of the substantive field under investigation. Technical literacy ensures operational accuracy, while disciplinary knowledge guides interpretation and prevents computational artefacts from being mistaken for meaningful patterns.

Kubsch considered this dual literacy essential for maintaining the human–machine dialogue that anchors CGT's reflexive integrity.

He also discussed some practical limitations. The method tends to perform less effectively with very short or homogeneous text segments, where models may detect superficial linguistic variations rather than substantive differences. Likewise, the representation of mathematical formulas in scientific discourse remains a challenge, highlighting the ongoing need for human interpretation. Nonetheless, even modest qualitative datasets can benefit from computational analysis when approached reflexively and transparently.

Despite these limitations, Kubsch views CGT as a flexible and scalable approach with considerable potential for science education research. Its transparency and reproducibility enable cumulative knowledge-building and collaborative validation, while its iterative structure allows adaptation to different data types and research goals. Even small qualitative datasets can benefit from it: *"Even if you have just ten interviews, each ten or twenty minutes, you can start using this."* For Kubsch, this flexibility, combined with its grounding in documentation and interpretive dialogue, makes CGT a powerful methodological resource for the field, a way to combine the analytical discipline of computational modelling with the sensitivity and contextual awareness that characterise qualitative inquiry.

5.6.3 Thematic Network Analysis with Dr. Jesper Bruun

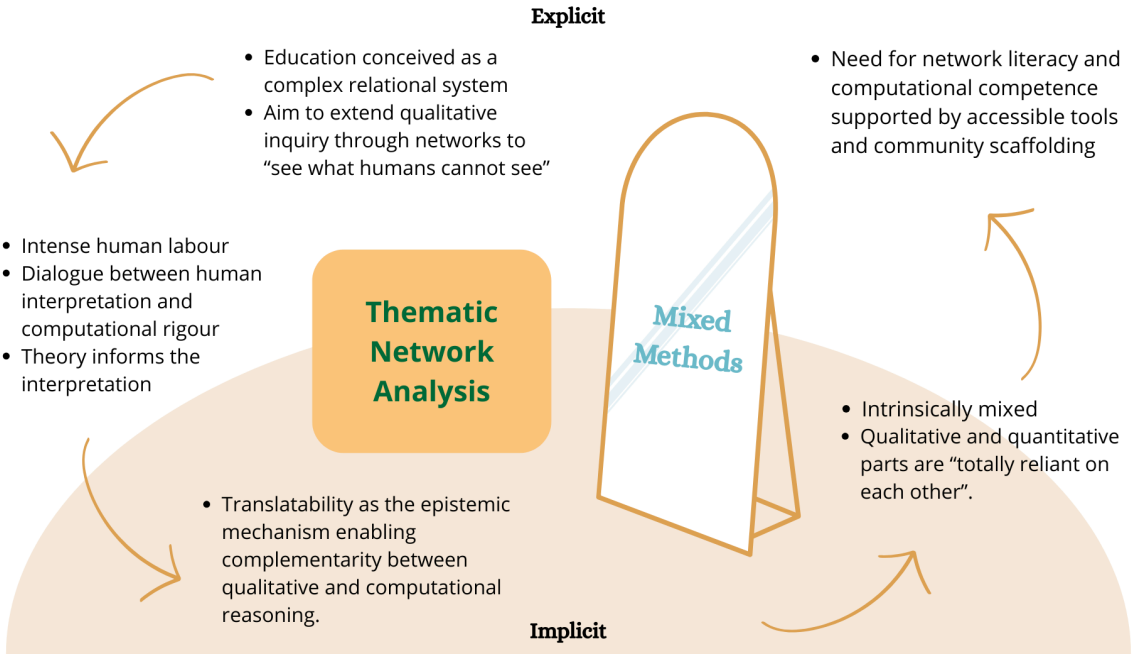

Figure 5.5 - Insights from the interview with Dr. Jesper Bruun about Thematic Network Analysis in Science Education Research

(a) Background, familiarity, and motivations for using the method

Jesper Bruun's trajectory as a researcher is marked by a persistent interest in representing the complexity of educational phenomena through analytical models capable of capturing relationships rather than isolated variables. His academic formation began in physics, where his doctoral research

already introduced a network perspective into educational contexts. As he recalled, *"I got to write a PhD in physics and it's called something about 'network in physics education research'. It was based on some work I had done with asking students who they could remember talking to in the preceding week, and we could then use those data to predict their future grades, and we could see sort of how some patterns emerged in those data"*. Through this early work, he developed both a conceptual and technical sensitivity to the relational structures that underlie learning processes. These experiences cultivated an enduring fascination with how complex systems (both social, linguistic or theoretical) can be analysed through network representations.

Bruun's orientation toward complexity was subsequently extended from the social to the linguistic domain. He described how his attention turned to language as a medium capable of capturing the relational and dynamic features of education: *"I had come across [the work of] some people called Masucci and Rogers who in 2006 had done linguistic network analysis where they had simply taken texts like Moby Dick and then converted that into a network and then they tried to make a model, but I was very intrigued with it: the way of thinking"*. This exposure to linguistic network analysis offered him a new way to conceptualise textual data: not as linear narratives but as interrelated webs of meaning.

From this interest, Bruun began to experiment with the use of network models in his own research. For example, *"From ESERA 2019, we did analysis of curricula across Europe. And so, I developed a method for making networks of such curriculum linguistic networks, where you could see different patterns emerge for different countries"*. These projects marked an important methodological transition. They consolidated his view that the study of education as an inherently complex and relational field that requires analytical approaches capable of capturing these interconnections.

Throughout the interview, Bruun expressed a distinctive passion for methodology itself: *"I really like methodology. So, I think I know something about most of the methods out there. I cannot say that I'm an expert in all of them, of course. But the reason for that is also [...] that I like to try and use network analysis with all the methods [...]. I think a lot of what we do in science education research is uncovering relations. And network analysis is about that. It is about relationships between entities"*. His words encapsulate an epistemic attitude characterised by openness and integration: a view of research methods as complementary tools for uncovering different facets of the same relational reality.

What ultimately convinced him of the method's potential was precisely this capacity to render visible phenomena that qualitative interpretation alone might overlook.*"It had sort of for me been a way of seeing things in the data that we as humans would not necessarily see",* he said. The motivation to integrate computational analysis into qualitative inquiry thus arose from a genuinely epistemic curiosity: a desire to extend, rather than replace, the interpretive capacity of the researcher. As he noted, *"When we started this study, we came in with the idea that maybe we wouldn't find anything, but maybe we would. And that's actually where I thought, 'OK, this method is something usable'."* From this perspective, TNA represented not a rupture but an expansion of qualitative inquiry, offering a concrete means to bridge human reasoning and computational modelling. For Bruun, such hybrid methods enhanced traditional qualitative analysis by combining interpretive depth with the systematic, scalable power of computation.

(b) Roles of the human researcher, the computational model, and theory

In describing how TNA operates in practice, Bruun portrayed the process as a constant negotiation between human judgment, computational precision, and theoretical insight. From the outset, the work

was characterised by a high degree of human involvement. The early stages required extensive manual curation and interpretive labour: linguistic cleaning, the grouping of synonyms, and the careful adjustment of textual rules to fit the specificity of the corpus. Such operations were not mechanical but interpretive, grounded in a shared understanding of meaning within context. Automated text-mining techniques available at the time, such as stemming or the removal of stop words, were deemed unsuitable for the small and conceptually rich corpus used in the study. The researcher's expertise was therefore indispensable in defining linguistic equivalences and exclusions that could preserve the semantic integrity of the texts.

Bruun emphasised that this interpretive phase was not merely preparatory but constitutive of the method itself. The decisions taken at this stage defined the representational space within which computational modelling would operate. He described the process as a continuous *conversation between the quantitative and the qualitative parts*, in which network analysis and discourse analysis iteratively informed one another. Certain expressions or concepts, such as "science education," which appeared with overwhelming frequency, had to be removed because their omnipresence obscured subtler relations. In other cases, language-specific features, such as the Swedish compound expression meaning "because," required preservation to maintain conceptual coherence. Each adjustment became an instance of translation between linguistic and computational representations, demonstrating that methodological rigour in TNA depends as much on interpretive attentiveness as on algorithmic precision.

The notion of reproducibility also acquired a distinctive meaning in this framework. Because every interpretive decision was recorded and could be re-run through the algorithm, the process itself became *traceable*. Bruun referred to this as a form of *inquiry audit*, a principle borrowed from qualitative research that gains new operational depth in computational contexts. The explicit documentation of choices allowed researchers to test variations, roll decisions back, and immediately visualise their implications. Through this recursive process, methodological reflexivity became embedded in the analytical workflow: human and algorithmic actions were not opposed but connected through a shared, revisable record of decisions.

An essential aspect of this dialogical configuration was the interplay between different kinds of expertise. Bruun worked alongside his co-authors Stine Mariegaard and Lars Seidelin. Stine, the first author, brought familiarity with early childhood science education that complemented his methodological competence in network analysis. Their collaboration exemplified a form of distributed cognition in which neither partner could independently assess the representativeness or interpretive adequacy of the results. As Bruun explained, his role was to identify structural patterns within the networks. These patterns were relations that appeared numerically or visually significant. The first author instead evaluated their educational and theoretical plausibility. The iterative exchange between these perspectives generated a methodological equilibrium: technical outputs became meaningful only through contextual interpretation, while theoretical insight gained precision and accountability through its translation into formal representations.

Theoretical knowledge played a guiding role in the analytical process. Pre-existing conceptual frameworks (such as Deweyan notions of learning by doing) shaped interpretive decisions in the identification and naming of modules. For example, disciplinary understanding informed choices about whether terms like "hands-on activities" and "student-centred experiments" could be treated as equivalent. In this sense, theoretical understanding was not external to the method but internal to its

functioning: it informed how linguistic expressions were grouped, how clusters were named, and how connections were judged to be meaningful.

In this sense, the human, the computational, and the theoretical did not occupy hierarchical roles but complementary ones. Human researchers supplied contextual understanding and interpretive discernment; the computational model ensured consistency, scalability, and procedural rigour; and theoretical knowledge mediated their dialogue, translating meaning into structure and structure back into meaning. The credibility of the analysis rested on the traceability of these choices, which made visible the dynamic movement through which human insight and algorithmic operation co-produced knowledge.

(c) Epistemological, ontological, and axiological assumptions

Bruun's reflections reveal a sophisticated and nuanced ensemble of epistemic and ontological assumptions underpinning TNA. His view departs from traditional realist conceptions of data as containers of pre-existing meanings waiting to be uncovered. When asked whether the patterns identified through TNA "already exist in the data" or are "constructed through the method", he responded unequivocally: *"I don't think it was there before… I think we are constructing something."* This statement expresses a constructivist epistemology in which knowledge does not emerge from discovery but from *methodological enactment*. The analytical process itself generates new forms of intelligibility by transforming texts into relational representations.

Yet this constructivism is not relativistic. Drawing on his background in physics, Bruun compared network models to theoretical constructs such as the Higgs boson: entities that are *"constructed"* yet nonetheless reveal aspects of reality that "suit a model". Knowledge, in this view, is model-dependent: networks do not mirror an external reality but provide a structured approximation that captures certain relational features of the phenomenon under study. The ontology implied here is therefore *model-based and relational*: reality is neither fully given nor entirely invented but *unfolds through representational practices* that connect theory, data, and computation.

Epistemically, this stance positions TNA as a framework for *relational knowledge construction*. The method's primary epistemic claim is not that it discovers universal truths but that it produces *relational coherence*: new insights emerge when dispersed linguistic or conceptual elements are connected through the dual action of algorithmic detection and human interpretation. As Bruun reflected on whether the patterns identified by TNA already exist in the data or are constructed through the method, he concluded that knowledge is produced through the act of modelling itself. *"I don't think it was there before,"* he said, *"but we are constructing something… it makes it more relational; new things are connected more."* In this sense, the networks do not simply display pre-existing relations but create a new representational form in which connections among ideas become visible and open to interpretation. The resulting map functions as a *representation of knowledge itself*—a structure that integrates the regularities detected by computation with the meanings derived from theoretical understanding

Validation, accordingly, is not sought through prediction or replication alone, although Bruun acknowledged that his scientific training inclines him to value predictive verification. When prediction is not possible, as in the analysis of theoretical discourse, validation occurs through *dialogical processes*: researchers return to the data, discuss interpretations collaboratively, and assess whether the identified relations are meaningful and credible. This approach substitutes *dialogical coherence* for statistical generalisation as the criterion of reliability. As he put it, "we talk to the other

authors and say, do you think this is there, and can we then go back into the data and find what we're looking for". In this epistemic model, the robustness of knowledge depends on *relational confirmation* rather than on independent replication.

These assumptions collectively articulate an epistemology grounded in translatability, that is the capacity for human, theoretical, and computational reasoning to be rendered into one another's terms. Translatability allows the method to integrate interpretive and formal operations without collapsing their differences. The human researcher formulates interpretive hypotheses and conceptual distinctions that are then translated into computational rules and parameters; the algorithm, in turn, generates new representations that can be re-interpreted within the theoretical framework. Knowledge arises through *iterations of translation*, each iteration refining both the model and the interpretation. This cyclical movement ensures that the dialogue between human and machine is not ancillary but constitutive of the knowledge-making process.

Ontologically, TNA embodies a view of reality as *relationally structured and representationally enacted*. What exists for inquiry are not isolated entities but networks of relations—semantic, conceptual, or theoretical—made visible through modelling. Epistemologically, it operates on the assumption that meaning and structure are co-produced: interpretive understanding informs modelling, and modelling reshapes interpretation. Axiologically, this process is sustained by values of reflexivity, transparency, and openness to revision. Bruun's insistence on documenting decisions and enabling iterative re-runs of the analysis translates these values into concrete methodological practices. Reflexivity here becomes procedural: the very architecture of the method ensures that analytical choices are explicit, revisable, and traceable.

Finally, Bruun's reflections also point to a deeper ethical dimension: *epistemic humility*. He recognised that network analysis may challenge established expectations of what scientific explanation should look like: *"it goes against what some people want from science education research"*. Yet rather than seeking closure or certainty, he advocated openness to methodological experimentation and to alternative ways of representing educational complexity. In his words, *"we have to be open about that… my experience right now is that it does offer some things that other methods could not offer."* This statement captures the spirit of the epistemological shift that methods like TNA embody: a movement from seeking universal generalisation to cultivating relational understanding, from prediction to reflexive iteration, and from static representation to dynamic translation.

(d) Positioning TNA in relation to mixed-methods research

In discussing the methodological status of TNA, Bruun expressed a clear and distinctive position: *"I think it really is [a mixed method], and I think it is more so than many others"*. His understanding of mixed methods differs markedly from conventional typologies in educational research, such as those proposed by Johnson and Onwuegbuzie (2007), which classify designs according to the predominance or sequencing of qualitative and quantitative components. These models, where one component often plays a supplementary or illustrative role, remain additive rather than integrative. As he put it, *"most of what I've seen is either the sort of big and small, where the qualitative part is maybe just a small set of interviews to inform you… Is that mixed? That is, in principle, mixed methods because you're using both, and they are sort of talking to each other."*

By contrast, TNA represents what he described as a *"true mixed-method approach"*. The two analytical strands are not juxtaposed but *"totally reliant on each other"*. *"You cannot say anything",*

he explained, *"unless you have both"*. The network analysis provides the structural grounding that makes thematic interpretation visible and traceable, while the qualitative interpretation supplies the contextual and theoretical meaning that renders the numerical relations significant. Without the networks, the claims would remain unsubstantiated; without the qualitative understanding, the networks would be "empty shells". This interdependence situates TNA as an intrinsically hybrid method, in which qualitative and computational operations exist in a state of reciprocal necessity rather than methodological hierarchy.

Bruun's view thus moves beyond the common understanding of mixed methods as the combination of data types or analytic techniques. For him, mixing occurs not *between datasets* but *between epistemic logics*, and extends across representational levels. Beneath the visual maps that make networks interpretable lies a mathematical apparatus, which gives formal coherence to the analysis. The visible network thus mediates between human interpretation and computational abstraction, making TNA a genuinely multi-layered and interdependent method, where each level (from visual to mathematical, from interpretive to formal) contributes to the same epistemic construction.

(e) Applications, competences, and potentials for Science Education Research

When reflecting on the present and future use of TNA in Science Education Research, Bruun identified both practical challenges and significant opportunities. He described the method as *labour intensive*, relying heavily on manual interpretation and iterative decision-making without risking the loss of semantic and contextual richness. Nonetheless, he expressed cautious optimism that advances in AI (particularly LLMs) could assist in these tasks, reducing manual effort while preserving interpretive transparency.

Beyond the technical demands, Bruun highlighted a deeper epistemic challenge: accessibility and understanding. Network analyses are still emerging within the field's methodological repertoire, and their interpretation requires forms of expertise that are not yet widely established. Network visualisations, for instance, demand new interpretive literacies, as they represent multiplicity and connectivity rather than the linear causality familiar from traditional statistics. Their difficulty arises not merely from complexity but from epistemic novelty, since TNA reveals relational phenomena that fall outside the expectations of generalisable, causal explanation. This shift, Bruun acknowledged, may unsettle researchers accustomed to linear models and effect sizes. Yet relational approaches such as TNA expand inquiry by making visible the heterogeneity of educational systems and the patterns that conventional methods obscure. He stressed, however, an attitude of openness: *Does this way of representing data model the phenomenon we want?* Such humility recognises TNA as one way of seeing, not a final truth, but one that renders previously inaccessible aspects of educational reality visible.

Looking ahead, Bruun envisioned the cultivation of new competencies enabling researchers to work effectively with relational and computational methods. Developing *network literacy* (akin to scientific or statistical literacy) requires both technical proficiency in building and manipulating networks and conceptual understanding of what these representations signify within educational theory. This must be complemented by computational competence and supportive infrastructures. Shared software environments, such as those emerging from the Epistemic Network Analysis community, can democratise access, allowing researchers from diverse backgrounds to engage meaningfully with complex analyses.

For Bruun, the broader significance of TNA lies in its capacity to bridge qualitative and quantitative traditions. By translating distinct methodological perspectives into a shared representational framework, it fosters dialogue and mutual understanding across the field. TNA thus represents more than a technical innovation: it is a methodological platform for cultivating new epistemic literacies and renewing the methodological imagination of science education research. Its future depends on the community's willingness to embrace relational thinking, develop interpretive and computational literacies, and sustain an open culture of translation between human and computational reasoning.

5.7 Discussion

5.7.1 Core ideas emerged from the analysis

The results emerging from this analysis are remarkably rich and multifaceted. Both the analytical work and, even more so, the three interviews with the authors of the selected studies have generated material of exceptional conceptual depth, opening reflections that extend far beyond what could be fully explored within this thesis. Their words offered not only technical clarifications but genuine methodological and philosophical insights into how AI and computational methods are reshaping the landscape of SER. I wish once again to express my sincere gratitude to Dr. Jesper Bruun, Dr. Peter Wulff, and Dr. Marcus Kubsch for their availability, intellectual generosity, and the depth with which they shared their experience.

The following sections present some key aspects that appear fundamental for understanding these emerging approaches in light of mixed-methods research and the broader methodological reflection they entail.

Ontology of language: reality as represented and representable

Across the cases analysed, language emerged as the primary medium through which reality is constructed, interpreted, and rendered accessible to analysis. In SER, particularly when studying reasoning, reflection, and learning, language is not merely a vehicle of thought but a representation of the phenomena themselves. Whether in students' essays, teachers' reflections, or scientific discourses, language constitutes a space where meanings, intentions, and relations materialise.

The computational approaches examined treat language as a structured system in which patterns can be detected and analysed. These patterns are not external to reality but expressions of it, revealing how scientific meanings are produced and circulated in both offline and online contexts. The ontology implied is thus semiotic and relational: the world of science education is approached through its linguistic manifestations, which can be modelled, visualised, and reinterpreted, also when large corpora are present. In this sense, language becomes both the object and the interface of inquiry, a domain where the qualitative and computational dimensions of research meet.

Iterativeness as epistemological dialogue

A second core insight concerns the iterative character of inquiry when computational methods are integrated with qualitative reasoning. Across all cases, knowledge developed through cycles of translation between interpretive and algorithmic actions: preparing the corpus, setting computational parameters, inspecting model outputs, and reinterpreting them theoretically. This continuous movement, between human reflection and computational representation, constitutes an epistemological dialogue, a conversation in which each act of analysis reformulates the previous one.

In this dialogue, qualitative reasoning becomes operationalised into algorithmic procedures, while computational representations re-enter the interpretive domain as new data to think with. Contextual understanding and numerical formalisation are not separable stages but alternating expressions of the same epistemic process. Each translation refines the phenomenon's representation, bringing to light both the potential and the limits of human and artificial intelligences.

Translatability as the epistemic mechanism of complementarity

This iterative exchange reveals the underlying condition that enables human and computational contributions to coexist within a single methodological configuration: translatability. Translatability can be defined as the capacity for interpretive reasoning and computational representation to be rendered into one another's terms through iterative reformulation. It constitutes the epistemic mechanism that allows complementarity, the mutual enhancement of human and machine intelligence in the production of knowledge.

Complementarity had already appeared empirically in the alternation between inductive and deductive logics, interpretive and algorithmic actions, theory-driven and data-driven reasoning. What the interviews make explicit is its *epistemic foundation*: the assumption that the output of one mode of reasoning can be translated into the language of the other without losing legitimacy. Translatability bridges two domains traditionally considered incommensurable, meaning and computation, by presupposing that interpretive constructs can be formalised, and that computational structures can be re-interpreted meaningfully within a theoretical frame. Meaning is not destroyed by formalisation; formal output is not devoid of meaning until interpreted.

In practice, this translation unfolds iteratively. The researcher converts interpretive choices into computational instructions (e.g. cleaning, grouping, parameterising) then re-encounters the model's output as a new representation of the phenomenon. Each iteration redefines the shared space where human understanding and algorithmic operation overlap. Knowledge is therefore not discovered but enacted through cycles of translation, each one leaving a trace that renders the reasoning transparent and reviewable. Reflexivity, in this framework, becomes operational: every choice is encoded in the analytical process and can be retraced, compared, and reinterpreted.

Translation, communication, and the generation of knowledge

This conception of translatability resonates with semiotic theories of communication, particularly Lotman's (1990) notion that *translation is the precondition for dialogue and the creation of new knowledge*. Communication, Lotman argues, is never neutral transfer but a process of negotiation between different worlds of meaning. For understanding to occur, each interlocutor must build a model of the other's perspective; comprehension requires difference and the effort of translation. In a "banal" communication, where sender and receiver already share a common code, no new knowledge arises. By contrast, in "non-banal" communication, meaning emerges precisely from the difficulty of translation, from the attempt to make intelligible what initially belongs to another semiotic world.

The hybrid methods examined in this thesis can be understood in similar terms. The dialogue between qualitative and computational reasoning represents a *non-banal communication* between two epistemic systems, each with its own vocabulary, logic, and mode of representation. Through the labour of translation (e.g. between theory and code, and between algorithmic output and human interpretation) new knowledge is generated. Without translation, there is no dialogue; without dialogue, there is no understanding.

Toward an epistemology of dialogue

Recognising translatability as the generative mechanism of complementarity implies an epistemological shift. Hybrid computational-qualitative methods do not merely combine techniques; they enact an *epistemology of dialogue*, where knowledge emerges through the circulation and transformation of representations between symbolic systems. Reality is not accessed directly but reconstructed through successive translations, each producing a partial yet communicable view of the phenomenon.

In this framework, the researcher's epistemic responsibility is to sustain communicability: ensuring that theoretical constructs can be formalised without distortion, and that algorithmic results can be re-contextualised without trivialisation. The credibility of knowledge no longer depends on the purity of paradigms but on the *traceability of translations*, that is the capacity to follow how meaning travels, changes, and stabilises across human and computational reasoning. Through this iterative dialogue, the mixed-methods tradition in science education expands into a genuinely relational and reflexive methodology, one that treats understanding itself as a product of translation.

5.7.2 Rethinking Mixed Methods and research paradigms

The empirical material analysed in this study provides an opportunity to reconsider both the scope of mixed-methods research and the broader landscape of research paradigms in science education. In particular, the methods examined, CGT, TNA, and ML and NLP approaches, invite us to rethink what methods can be mixed, how they are combined, and to what extent paradigms themselves may need to expand to accommodate emerging practices.

Rethinking what makes these studies "mixed"

Each of the three cases analysed exhibits a distinct form of methodological integration. Some, such as Thematic Network Analysis or NLP-based approaches, can be characterised as *inherently mixed methods*, meaning that mixing is embedded at every stage (such as conceptual, analytical, and interpretive) making any separation between strands impossible. This view implies a paradigmatic rather than merely procedural understanding of integration. Others, like Computational Grounded Theory and certain ML applications, can be described as *hybrid methods*, frameworks in which computational procedures are embedded within, or developed alongside, established qualitative traditions.

What makes these studies mixed is not simply the co-presence of qualitative and quantitative techniques, but the integration of *different logics of inquiry*: interpretive reasoning and computational formalisation. They merge the sensibility of qualitative inquiry, that includes contextual sensitivity, theoretical grounding, and interpretive richness, with the systematic precision and scalability characteristic of computational analysis. Yet they do so in ways that differ substantially from traditional conceptions of mixed methods.

Beyond the classical model of integration

First, these studies do not rely on two distinct datasets of different natures, as is common in pragmatic mixed-methods designs. Instead, they work with a *single dataset*, typically textual, subjected to both qualitative interpretation and computational transformation (for example, through embeddings, distance metrics, frequency counts, or network modelling). This move challenges the conventional

expectation that mixed methods require multiple forms of data; instead, it shows that methodological integration can occur within one dataset, through the interplay of distinct analytical logics.

Second, the relationship between methods is not additive or sequential but *iterative and dialogical*. Human researchers and computational systems engage in a continuous cycle of decisions, refinements, and interpretations, each informing the other in real time. In several cases, such as Thematic Network Analysis, the two "sides" of the method cannot be meaningfully separated: network analysis and thematic interpretation operate as co-dependent dimensions of the same epistemic process.

Third, the balance of roles and values reveals a distinctive form of complementarity. Human agency is not erased by computational automation, yet interpretation is not positioned as inherently superior to quantification. Rather, these studies highlight the *orchestration of heterogeneous epistemic resources*: researchers contribute contextual awareness, domain knowledge, and critical judgement, while computational methods offer reproducibility, scalability, and the ability to reveal latent relational structures. The outcome is not a pragmatic merger of two traditions but the emergence of a *new generation of mixed methods*, in which integration is embedded in the very architecture of analysis.

Revisiting research paradigms

The originality of these approaches also prompts a reconsideration of paradigms themselves. Drawing on the framework proposed by Treagust and Won, three possible directions can be identified.

1. **Extension of the post-positivist paradigm:** Some studies explicitly frame their methods as quantitative, even when applied to textual data. Network analysis, for instance, is defined by its authors as a quantitative technique, grounded in measurement, reproducibility, and prediction. From this perspective, computationally enhanced text analysis can be seen as *extending post-positivism*, by expanding the range of phenomena, textual and qualitative in origin, that can be rendered measurable and statistically tractable.
2. **Extension of the interpretivist paradigm:** At the same time, many of these methods are rooted in qualitative traditions, with computational tools introduced as *extensions or enhancements* to overcome the limitations of traditional approaches. The emphasis remains on interpretation, theoretical sensitivity, and contextual awareness. However, it would be reductive to regard these studies as purely interpretivist. Here, the human researcher is not a figure of epistemic superiority but a *mediator* orchestrating diverse sources of evidence and meaning. Computational procedures expand, rather than diminish, the interpretive act.
3. **Towards an extension of the mixed-methods paradigm:** The position advanced in this thesis is that these practices are best understood as extending the very idea of mixed methods itself. Traditional mixed-methods research has been defined by the integration of qualitative and quantitative approaches. What emerges here is a *second genre* of mixed methods, in which computational and AI-based approaches enter as a *third epistemic pillar* alongside qualitative and quantitative. This suggests envisioning a triadic paradigm in which mixed-methods research is conceptualised not only as *qualitative-quantitative* combinations, but also as *qualitative-computational* (as demonstrated in this thesis) and potentially *quantitative-computational*.

Such a move inevitably raises foundational questions. What are the ontological, epistemological, and axiological assumptions that distinguish computational/AI methods from traditional quantitative and qualitative approaches? How should we conceptualise the tensions and synergies among these three

pillars, qualitative, quantitative, and computational? Within this thesis, some of these questions were addressed but are not yet resolved. For sure, this condition points toward an evolving research landscape in which AI becomes recognised as a legitimate paradigm of inquiry in its own right.

This ongoing transformation is not accidental. It occurs at a historical moment when AI itself is emerging as a new epistemic agent, prompting redefinitions across disciplines. Even within the philosophy of AI, scholars such as Ganascia have noted its *dual nature*, at once symbolic and empirical, interpretive and computational, positioning AI as a bridge between knowledge traditions.

The analyses presented in this thesis suggest that computational-qualitative integration is not a marginal or incidental practice, but a robust methodological form with the potential to reshape how research designs and paradigms are conceived in science education. By embracing these emerging forms, mixed-methods research can evolve beyond its dualistic heritage into a genuinely pluralistic epistemology, capable of sustaining dialogue among diverse ways of knowing in the age of AI and data science.

5.8. Conclusions

Through the analysis of exemplar studies in science education research and the expert interviews guided by the heuristic tool, this chapter has examined in depth how computational methods are employed within qualitative inquiry. This exploration allowed us to address a central question that has accompanied the entire trajectory of this doctoral work: how can we integrate two forms of knowledge that originate from distinct worldviews—the post-positivist and calculable on one side, and the interpretivist and qualitative on the other?
The analyses revealed that such integration becomes possible through what we have termed the *mechanism of translation*: the iterative process through which human researchers and computational systems exchange representations, each rendering the other's outputs intelligible and meaningful. This mechanism makes complementarity between human interpretation and computational formalisation not only possible but epistemically productive. It is precisely through this dynamic of translation that new *inherently* and *hybrid* mixed methods emerge—approaches in which the boundaries between paradigms are not erased, but rendered permeable, giving rise to genuinely dialogical forms of methodological inquiry.

References

Blei, D. (2011). Introduction to Probabilistic Topic Models. *Communications of the ACM, 55*, 77-84. https://doi.org/10.1145/2133806.2133826

Castillo-Montoya, M. (2016). Preparing for Interview Research: The Interview Protocol Refinement Framework. *The Qualitative Report*, 21(5), 811-831.

Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., Slater, S., Baker, R., & Warschauer, M. (2020). Mining Big Data in Education: Affordances and Challenges. *Review of Research in Education*, *44*(1), 130–160. https://doi.org/10.3102/0091732X20903304

Ganascia, J. G. (2010). Epistemology of AI Revisited in the Light of the Philosophy of Information. *Knowledge, Technology & Policy,* 23, 57–73. https://doi.org/10.1007/s12130-010-9101-0

Glaser, B., & Strauss, A. (1967). The Discovery of Grounded Theory: Strategies for Qualitative Research. Mill Valley, CA: Sociology Press.

Lotman, J. (1990). The Universe of the Mind: The Semiotic Theory of Culture. I.B. Thauris & Ltd.

Mariegaard, S., Seidelin, L. D., & Bruun, J. (2022). Identification of positions in literature using thematic network analysis: the case of early childhood inquiry-based science education. *International Journal of Research & Method in Education*, 45(5), 518-534.

Nelson, L. K. (2020). Computational Grounded Theory: A Methodological Framework. *Sociological Methods & Research*, 49(1), 3–42. https://doi.org/10.1177/0049124117729703

Rosenberg, J. M., & Krist, C. (2021). Combining Machine Learning and Qualitative Methods to Elaborate Students' Ideas About the Generality of their Model-Based Explanations. Journal of Science Education and Technology, 30(2), 255–267. https://doi.org/10.1007/s10956-020-09862-4

Tschisgale, P., Wulff, P., & Kubsch, M. (2023). Integrating artificial intelligence-based methods into qualitative research in physics education research: A case for computational grounded theory. *Physical Review Physics Education Research,* 19, 2. https://doi.org/10.1103/PhysRevPhysEducRes.19.020123

Vollstedt, M., & Rezat, S. (2019). An Introduction to Grounded Theory with a Special Focus on Axial Coding and the Coding Paradigm. In: Kaiser, G., Presmeg, N. (eds) Compendium for Early Career Researchers in Mathematics Education. ICME-13 Monographs. Springer, Cham. https://doi.org/10.1007/978-3-030-15636-7_4

Willig, C. (2013). Introducing Qualitative Research in Psychology. McGraw-Hill Education.

Wulff, P., Buschhüter, D., Westphal, A., Mientus, L., Nowak, A., & Borowski, A. (2022). Bridging the Gap Between Qualitative and Quantitative Assessment in Science Education Research with Machine Learning—A Case for Pretrained Language Models-Based Clustering. *Journal of Science Education and Technology*, 31(4), 490–513. https://doi.org/10.1007/s10956-022-09969-w

Wulff P., Westphal, A., Mientus, L., Nowak, A., & Borowski, A. (2023). Enhancing writing analytics in science education research with machine learning and natural language processing—Formative assessment of science and non-science preservice teachers' written reflections. *Frontiers in Education*, 7. https://doi.org/10.3389/feduc.2022.1061461

General conclusions of the thesis

This thesis set out to explore how the emergence of Artificial Intelligence (AI) and computational methods is reshaping the methodological foundations of Science Education Research (SER), with particular attention to the analysis of textual data. What began as an inquiry into how to integrate qualitative sensitivity with computational breadth evolved into a reflection on the very nature of method—its assumptions, values, and epistemic ambitions. The work has argued that, in the age of AI and big data, methods are not neutral technical devices but active mediators of knowledge, carrying within them the philosophical and cultural traces of the paradigms from which they arise.

The thesis first situated the study within the broader field of mixed methods research, showing how traditional boundaries between qualitative and quantitative paradigms are increasingly porous. The growing presence of AI challenges these boundaries, giving rise to hybrid epistemologies where interpretation and computation coexist within the same research act. By processing vast amounts of data while generating interpretive structures, AI enables new forms of methodological pluralism in which positivist precision and interpretive depth no longer stand in opposition but interact within an integrative and dynamic methodological landscape.

This theoretical argument was grounded in two empirical explorations. The first case study applied unsupervised natural language processing to a large corpus of physics education literature, revealing long-term thematic trends and showing how computational methods can map the intellectual evolution of a field. The second analysed a medium-sized corpus of students' essays through both computational and qualitative lenses, demonstrating that when data-driven discovery meets interpretive reasoning, new insights emerge. Together, these studies highlighted both the potential and the challenges of combining computational and qualitative perspectives in educational text analysis.

From these experiences emerged a crucial recognition: mastering a method is not enough—understanding what lies beneath it is essential. This insight led to the development of a heuristic tool designed to make visible the ontological, epistemological, and axiological assumptions embedded in research methods. Conceived as both an analytical framework and a reflective device, the heuristic enables researchers to examine how data are defined, how knowledge is constructed, and which values underpin methodological choices. It thus offers a structured way to navigate the expanding methodological landscape and to situate new approaches critically within it.

The heuristic was subsequently applied in two directions. First, it was used to interrogate AI-based methodologies in scientific research, revealing a profound redefinition of what counts as a "method." AI emerges not merely as a technical instrument but as an epistemic agent capable of generating, transforming, and curating data in ways that reshape scientific reasoning itself. Second, the heuristic was used to explore the integration of computational techniques within Science Education Research, combining the analysis of recent studies with expert interviews. This investigation revealed novel, hybrid forms of mixed methods—dynamic exchanges between quantitative modelling and qualitative interpretation—that challenge traditional procedural definitions of "mixing."

Across these analyses, a central argument emerged: methodological innovation in the age of AI must be accompanied by methodological awareness. The expansion of computational possibilities should not lead to the abandonment of established paradigms, but to their re-examination from within. Each method—whether human or algorithmic—embodies a way of seeing and knowing the world.

AI-based approaches thus invite not only technical proficiency but epistemic literacy: the capacity to understand, evaluate, and teach methods as culturally and philosophically situated practices.

In this sense, the thesis contributes to a redefinition of mixed methods as a paradigmatic rather than procedural concept—one that recognises the coexistence of multiple ontologies, epistemologies, and axiologies within research practice. It also calls for a renewed dialogue between science education and the philosophy of science, recognising that the methods we use to study learning are themselves forms of learning about the world.

This work contributed to rethinking methodological foundations from both philosophical and practical perspectives, and to revisiting qualitative, quantitative, and mixed methods in Science Education Research in the age of AI and data science. Through the development and application of the heuristic tool, and the analyses conducted throughout this thesis, it became possible to articulate conceptual and methodological language capable of describing these ongoing transformations. More importantly, this work contributed to clarifying, in practice, the conditions under which these methods operate, the domains within which their claims are valid, and the criteria by which their epistemic contributions can be meaningfully evaluated.

Understanding more deeply how qualitative approaches and computational methods can be integrated—and through which mechanisms this integration becomes epistemically productive—has allowed this work to illuminate pathways for developing transparent and fruitful methodologies that meaningfully embed AI within Science Education Research. In a world increasingly shaped by AI and technological transformation, researchers and educators alike must not remain passive recipients of these changes but active participants in shaping them. Building methodological awareness, fostering dialogue between human interpretation and computational modelling, and cultivating reflexivity in the use of AI are essential steps toward ensuring that future research and education remain not only rigorous and innovative but also profoundly human.

Acknowledgments

Even if I am not fully able to realise it yet, the moment of writing these final lines has arrived. What I am experiencing now is certainly a profound sense of pride and surprise. Yet this beautiful feeling is accompanied by flashes of restlessness and fear. Reflecting on it, I realise that this contrast is not unexpected, as the entire PhD journey has been shaped by such diverse feelings, in the most fascinating sense. It has been a constant coexistence of intense and opposing emotions: effort and excitement, curiosity and fear, receiving support and offering it to others, tears and moments of paralysis, beers to celebrate achievements and glasses of water to refresh tired thoughts, feelings of being lost and epiphanies when many scattered pieces suddenly aligned. I knew this would be an extraordinary adventure, both in its difficulties and in its joys, but I never imagined it would become the hardest and richest experience of my life.

This *tension*, as we often called it, is precisely what drew me to the possibility of pursuing a PhD. It allowed me to inhabit a *boundary zone* between learning, research, and teaching; to expand my knowledge by immersing myself in what was unknown; and, perhaps most importantly, to learn how to stay with uncertainty rather than eliminate it, first and foremost as a human being.

The gratitude I wish to express in these lines is therefore amplified, knowing how much complexity, growth, and transformation this journey has woven together.

First, I would like to sincerely thank Professor Marcus Kubsch and Professor David Treagust for agreeing to read and evaluate this dissertation. Your careful reading, thoughtful feedback, and constructive suggestions helped me to consolidate and refine this work, and I am deeply grateful for the attention and care you dedicated to it.

I wish to express my deepest gratitude to my supervisor, Professor Olivia Levrini. You have been the person who experienced this tension most closely with me, offering constant and generous support at every stage of this journey. I had the privilege of working with someone who consistently expressed trust and confidence in my abilities, even at times when I struggled to do so myself. You taught me many ways of inhabiting this path: how to cultivate refined and heartfelt thinking, how to look ahead where no one is looking, how to “fail fast”, how to search for solutions, how to transform ideas into action, how to embrace complexity, and how to trust the research process. You also taught me how to have both “big eyes” and “small eyes”: to hold together vision and attention to detail. For all of this, and much more, I am profoundly grateful.

I would also like to thank my co-supervisor, Professor Tor Ole B. Odden. Everything began with a clear and structured Google search. The keywords I entered led me to your work, and then to you, and eventually, to Oslo. That encounter, and our subsequent collaboration, has been one of the most meaningful and unexpected gifts of these four years. You are an exceptionally rigorous and thoughtful researcher, and I learned a great deal from your way of thinking, questioning, and engaging with research. Your supervision enriched this work and strengthened my confidence in becoming a researcher myself.

I am also deeply grateful to all the people I met at the Center for Computing in Science Education at the University of Oslo. Thank you for the stimulating discussions, the exchange of ideas and cultures, and the companionship during my time there. And I would like to thank the city of Oslo itself, to which I will always feel connected and grateful. Even recalling it now evokes a profound sense of

peace, inspired by its breathtaking nature, and a feeling of home, for having welcomed me for some months of my life in its endless light, truly an embrace for the soul.

I would like to thank all the researchers and colleagues who, in different ways, enriched my path and supported both me and my research, making me feel part of an extraordinary research group.

In particular, I thank Giulia Tasquier, Sara Satanassi, Paola Fantini, and Laura Branchetti, for being constant points of reference for the many needs a PhD student encounters, and for always being ready to offer thoughtful feedback, support, and encouragement. I also thank Sibel Erduran for being a reference point and an inspiring person every time we had the opportunity to meet and collaborate.

I thank Lorenzo Miani and Francesco De Zuani Cassina, who started this adventure with me and were companions in countless hours of discussion, exchange, lunches, and shared moments throughout this PhD. Being together was a key element of this journey.

I thank Emma D'Orto for joining our research group and being someone with whom I deeply resonated, bringing joy, understanding my most difficult moments, and supporting me like a sister.

I thank Eleonora Barelli and Andrea Zanellati, who were important points of reference in this fragmented and unique PhD in Data Science and Computation. At many times, you embodied what I aspired to become, both as researchers and as people.

I thank all those who, year by year, joined our research group, in particular Francesca Riccioni, Saul Casalone, Veronica Ilari, and Maria Teresa Scarongella, for the exchanges, conversations, and shared moments.

All of them, together with the many visiting researchers who spent time in our group and became friends, contributed to making our research group a group I am proud to be part of, showing me many creative research paths, always starting from the same beating heart for research in physics education.

My sincere thanks also go to the PhD coordinators, the members of the doctoral committee, and my colleague Alessandra Stramiglio, with whom I shared the role of student representative. I am also grateful to the European Science Education Research Association community, which provided spaces for intellectual exchange, inspiring ideas, and a sense of belonging.

I would like to thank my family. To my parents, Alberto and Marinella, for their constant support and trust in me, even when I could not fully explain "what I was doing at university". Your faith in me has meant everything. To my brother Davide, and Miriam, who became family to me as well, and to my sister Daniela, for your help and for making me feel free to be completely myself. To my cousin and dear friend Ilaria, for being proud of me and showing it, for checking on me wherever I was, and for growing alongside me. To my big family, my grandmother Bianca, my aunts, my uncles, and my cousins. I felt your interest, encouragement, and affection throughout this journey. I also thank Ettore and Carla, for their enormous generosity and for welcoming me into their family, and Erica and Mattia, for their presence, dialogue, and shared moments of joy.

I am deeply grateful to my friends. To Chiara, who has been beside me since our studies and throughout this doctoral journey: My PhD wouldn't have been the same without you by my side. To Alice, Arianna, and Alessia, for their friendship and closeness. To Andreina and Alice, long-time friends and fellow PhD travelers, for understanding what it means to walk the PhD path and for always being there to lift me up, even from miles and miles away. To my friends from Reggiolo and

Guastalla. To Paola and Martina, my housemates in Bologna, who welcomed me and made our time together fun, meaningful, and peaceful.

To Mirco, my love. Every day, together, we have built something new, with patience, care, and dedication. You have been present through every moment of this journey, offering love, strength, and quiet support. Your presence gave me stability when everything felt uncertain, and your faith gave me courage when I needed it most. Thank you for walking beside me in this chapter of life, and in those still to come. I love you so much.

San Bernardino, Novellara
15th February 2026

Appendix A

AI-based methods in science mapped from the *Nature* article *"Scientific discovery in the age of artificial intelligence"* (Wang et al., 2023).

Scientific function	AI technique / category	Specific method	Description	Examples /use in science
Data selection	Deep Learning (Unsupervised)	Anomaly detection (e.g., Autoencoders)	Identifies rare or unexpected events by measuring reconstruction error from background patterns	Detection of rare astronomical phenomena; medical diagnostics
Data annotation	Supervised Machine Learning	Pseudo-labelling	Uses a model trained on a small labelled dataset to annotate large unlabelled datasets	Ecological monitoring; image and video datasets
	Graph-based ML	Label propagation via similarity graphs	Propagates known labels across graph nodes based on feature similarity	Predicting molecular properties; chemical classification
Data selection + annotation	ML + Bayesian Statistics	Active learning (Bayesian optimization)	Improves model accuracy by selecting the most informative data points for human annotation	Genomic studies; materials science
Data generation	Generative AI	Automatic data augmentation (Reinforcement Learning)	Learns a policy to automatically augment datasets for diversity or coverage	Scientific images, molecular design
		Deep Generative Models (VAEs, GANs)	Learns the underlying data distribution to generate new synthetic samples	Images or sequences in astronomy, high-energy physics, biology
Data refinement	Convolutional DL / Autoencoders	Denoising and resolution enhancement	Removes noise and enhances resolution in image and sequence data	Live cell imaging; black hole imaging; RNA-protein expression analysis

Representation learning	Deep Learning (Self-supervised)	Feature learning from unlabelled data	Learns useful patterns or features without requiring manual labels (pre-training step)	Satellite imagery; cell imaging
	NLP + Deep Learning	Masked language modelling (e.g., Transformer models)	Predicts masked elements in sequences using attention-based neural networks	DNA and protein sequence modelling; earthquake signal detection
	Geometric DL + Graph ML	Geometric representation learning	Learns embeddings that reflect structural or spatial relationships encoded in graphs	Protein folding (AlphaFold); particle physics
Hypothesis generation from data	Symbolic regression + Reinforcement Learning	RL-generated symbolic expressions	Learns policies to sequentially construct mathematical formulas by selecting notation symbols; uses rewards to evaluate formula quality	Discovering symbolic laws in physics; mathematical conjecture refutation
	Evolutionary algorithms	Symbolic law evolution	Generates symbolic expressions via mutation and selection over generations to fit observational data	Discovery of empirical formulas from data
Hypothesis search in combinatorial spaces	Reinforcement Learning	Molecule generation with policy learning	Learns how to build molecules step-by-step by choosing atoms and positions to optimize chemical properties	Drug discovery; protein design
Hypothesis space navigation	Active Learning + Bayesian Optimization	Iterative candidate refinement	Selects most informative or promising hypotheses to test next based on uncertainty or expected improvement	Ligand screening; particle physics; protein function prediction

Hypothesis filtering	Weakly supervised learning	Black-box predictors with noisy labels	Trains models to predict hypothesis quality using noisy, indirect, or simulated supervision when experimental data are limited	Molecular prioritization in drug screening
Learning hypothesis distributions	Probabilistic ML	Bayesian posterior over hypotheses	Learns full distributions over plausible hypotheses compatible with prior knowledge and observed data	Symbolic expression sampling; candidate generation under uncertainty
Hypothesis optimization	Deep Generative Models (VAEs, Grammar VAEs)	Latent space optimization	Maps hypotheses (e.g. formulas, molecules) to continuous latent space for gradient-based optimization, then decodes optimal candidates	Protein design, symbolic law discovery, gravitational wave detection
	Differentiable Programming	Sparse symbolic regression	Uses trainable basis functions and regression to identify compact, interpretable laws from data	Discovering dynamic systems' rules in physics
	Deep Generative Models (e.g. SMILES-VAE)	Molecule optimization in latent space	Encodes molecules into differentiable space; backpropagates gradients to optimize for properties, then decodes to discrete structures	Drug discovery, protein folding
Hypothesis generation	LLMs and Transformers	Hypothesis suggestion from text or sequences	Models generate hypotheses (e.g., biological variants or molecular candidates) directly from data or literature	Genomics (DNA variant prioritization); symbolic conjectures
Hypothesis discovery	Representation learning + symmetry detection	Symmetry-minimizing loss function	Learns smooth loss to identify symmetries as emergent properties, e.g. in physical systems	Discovering hidden structures in black hole data

Experimental planning	Active Learning	Uncertainty-guided experimental design	Selects most informative experiments based on estimated uncertainty, reducing total required tests	Materials synthesis; chemical space exploration
	AI-guided synthesis planning	Retrosynthesis prediction	Predicts stepwise synthesis routes for a target molecule from available precursors	Organic chemistry synthesis planning
Experimental steering	Reinforcement Learning	Policy learning for dynamic control	Adapts in real-time to experimental feedback, optimizing process control under uncertainty	Tokamak plasma control; quantum systems; stratospheric balloon navigation
Experimental optimization	Reinforcement Learning	Closed-loop experiment optimization	Interacts with experimental environment to maximize efficiency/success via feedback	Quantum system control; dynamic experimental design
Simulation enhancement	Deep Neural Potentials	Learned force fields	Replaces manual molecular force fields with AI-learned potentials from quantum-mechanical data	Molecular dynamics simulations
Simulation acceleration	Coarse-graining + ML	Learned abstraction for large systems	Learns to simplify large molecular systems without losing critical behavior	Coarse-grained molecular dynamics
Simulation inference	Neural PDE solvers	Physics-informed neural networks	Solves PDEs by combining physical laws with deep learning constraints	Fluid dynamics; chemical kinetics; seismic wave modeling
	Neural ODEs	Continuous-time modeling	Models system evolution over time using learned differential equations	Spatiotemporal system dynamics

	Operator learning	Deep operator networks	Learns mappings between functions rather than inputs and outputs for solving complex equations	Navier–Stokes solvers; PDEs with boundary conditions
Simulation adaptation	GNN + Neural PDEs	Graph-discretized solver adaptation	Adapts PDE solvers to irregular or unstructured domains using graph partitioning	Complex domain fluid simulations
Statistical simulation	Deep Generative Models	Boltzmann generators (normalizing flows)	Models full distributions over equilibrium states using invertible neural nets	Lattice field theory; statistical physics simulations
Statistical modeling	Deep generative models	Density estimation via normalizing flows	Learns high-dimensional probability distributions for system states	Sampling in Markov Chain Monte Carlo; quantum field theory