

Knowing with the Machine: Towards an Epistemology of Research Augmented by Large Language Models in the Social Sciences

Rachid Maghniwi

Phd, LARMODADD Laboratory, Faculty of Economics, Legal and Social Sciences Souissi, Mohammed V University of Rabat, Morocco

ABSTRACT

The entry of large language models (LLMs) into social science research practices, such as interview coding, literature reviews, hypothesis generation and the simulation of respondents, is usually discussed from a technical or ethical standpoint. This paper offers an epistemological reading. It argues that an LLM is neither a neutral instrument nor a "co-researcher", but an interpretive mediator whose probabilistic and opaque outputs affect the criteria of rigour differently depending on the paradigm adopted. By confronting LLM uses with the positivist, critical realist, interpretivist and constructivist paradigms, the paper shows that each paradigm calls for specific safeguards. It then proposes a protocol for comparing human and algorithmic coding, followed by a five-dimension rigour framework (traceability, reproducibility, reflexivity, adversarial validation and interpretive accountability) specified for each paradigm. The contribution is twofold. Theoretically, the paper brings generative AI into the debate on the foundations of knowledge in the social sciences. Methodologically, it provides researchers and reviewers with a tool to assess the legitimacy of AI-assisted research.

Keywords: epistemology, large language models, generative artificial intelligence, research methodology, validity, reflexivity, social sciences, epistemological positioning, research rigour

1. INTRODUCTION

Since the mass diffusion of conversational agents based on large language models (LLMs) at the end of 2022, social science research practices have been undergoing a rapid transformation. Researchers now use these tools to annotate text corpora, summarise literature, propose analytical categories, translate verbatim material and draft data collection instruments. Studies have shown that these models can outperform crowd-sourced annotators on certain text classification tasks [1]. Others propose using them to simulate samples of human respondents, the so-called "silicon samples" [2]. Still others explore their use in qualitative thematic analysis [3] or across the full range of computational social science tasks [4]. This enthusiasm has been accompanied by warnings. LLMs have been described as systems that assemble plausible linguistic sequences without reference to meaning [5]. Messeri and Crockett [6], for their part, point out that AI can create illusions of understanding among researchers: the feeling of understanding more than one actually does, or of holding knowledge that is more objective than it really is. These criticisms, however, mostly concern technical reliability or research ethics. They leave open a more

fundamental question: **what happens to the validity criteria of knowledge when a probabilistic system takes part in producing it?**

This question belongs to epistemology, understood as the study of how valid knowledge is constituted [7]. In management science, as in the social sciences more broadly, the French-speaking tradition gives a central place to the researcher's "epistemological positioning" [8, 9]. This positioning specifies the nature of the knowledge sought, the way it is developed and the criteria by which its value is judged. It would therefore be inconsistent to integrate a tool as powerful as an LLM into the research process without examining its compatibility with the paradigm the researcher claims to follow.

The paper starts from an observation: the use of LLMs is currently treated as a matter of tooling, whereas it concerns the very status of the knowledge produced. It formulates the following research question: **to what extent does the use of LLMs alter the conditions of validity of knowledge in the social sciences, and how do these conditions vary according to the epistemological paradigm adopted?** Three sub-questions follow:

1. What epistemological status should be granted to the LLM in the research process?
2. Which criteria of rigour are weakened, or on the contrary strengthened, in each paradigm?
3. Which methodological safeguards can preserve the legitimacy of AI-assisted research?

To answer these questions, the paper adopts a conceptual approach, complemented by the proposal of an empirical protocol comparing human coding and LLM coding. Section 2 reviews the main epistemological paradigms and their validity criteria. Section 3 analyses the LLM as an interpretive mediator and examines its effects paradigm by paradigm. Section 4 presents the comparison protocol. Section 5 proposes a rigour framework for AI-assisted research. Section 6 discusses the contributions, limitations and avenues for future research.

2. THEORETICAL FRAMEWORK: EPISTEMOLOGICAL PARADIGMS AND VALIDITY CRITERIA

2.1. Epistemological Positioning as a Requirement for Coherence

Every research project rests, explicitly or not, on answers to three questions: what is the status of the knowledge produced, how is it generated, and how can its value be justified [10]. The first involves an assumption about the nature of the reality under study (the ontological dimension). The second concerns the ways of accessing this reality (the methodological dimension). The third relates to the criteria that make it possible to say that a piece of knowledge is valid (the epistemic dimension in the strict sense).

The typology of Burrell and Morgan [11] long structured these choices around two axes: an objective or subjective conception of social science, and an orientation towards regulation or towards radical change. The French-speaking management literature has since refined this map, distinguishing in particular between positivism, critical realism, interpretivism and two forms of constructivism [8, 9, 12]. The aim is not to rank these paradigms, but to require coherence between the paradigm claimed, the question asked, the methods used and the evaluation criteria adopted.

2.2. Four Reference Paradigms

Positivism and post-positivism. Social reality is conceived as independent of the observer and governed by regularities that research can discover. Knowledge progresses through the formulation of hypotheses that are tested against the facts. Post-positivism, inspired by Popper [13], abandons the ideal of definitive verification in favour of falsifiability: knowledge is provisionally retained as long as it withstands attempts at refutation.

Critical realism. Stemming from the work of Bhaskar [14], critical realism posits a reality stratified into three domains: the real (generative structures and mechanisms), the actual (the events produced by these mechanisms) and the empirical (what is observed). Observed regularities are not sufficient: research seeks to identify, through retroductive reasoning, the mechanisms that explain them [15].

Interpretivism. Social reality is made up of the meanings that actors attribute to their actions. To know is to understand these meanings in their context. Knowledge is situated and depends on the interaction between the researcher and the actors studied [16].

Pragmatic constructivism. In the tradition of Le Moigne [10] and Avenier [12], this paradigm takes no position on the existence of an independent reality, but holds that one can only know one's experience of it. Knowledge is a construction whose value lies in its viability, that is, in its capacity to guide action relevantly in a given context.

2.3. Validity Criteria Specific to Each Paradigm

Each paradigm defines its own criteria of rigour. The positivist paradigm favours internal validity, external validity, reliability and objectivity. Lincoln and Guba [17] proposed parallel criteria for naturalistic inquiry: credibility, transferability, dependability and confirmability. Pragmatic constructivism emphasises the functional fit of knowledge with experience and the explicit justification of each inference [9]. Finally, critical realism judges knowledge by its explanatory power, that is, by its ability to account for the mechanisms at work.

Table 1. Epistemological Paradigms and Associated Validity Criteria

Paradigm	Status of knowledge	Dominant mode of reasoning	Validity criteria
Positivism / post-positivism	Knowledge of an independent reality, provisionally corroborated	Deductive, hypothetico-deductive	Internal and external validity, reliability, objectivity, replicability
Critical realism	Fallible knowledge of generative mechanisms	Retroductive and abductive	Explanatory power, plausibility of mechanisms, triangulation
Interpretivism	Understanding of situated meanings	Inductive, hermeneutic	Credibility, transferability, dependability, confirmability
Pragmatic constructivism	Viable representation of experience	Abductive, iterative	Functional fit, justification of inferences, actionability

Source: Author, based on Lincoln and Guba [17], Allard-Poesi and Perret [8] and Avenier and Thomas [9]

This table provides the analytical lens for the rest of the paper. If an LLM takes part in the production of knowledge, one must ask which criteria in the last column are affected, and in what way.

3. THE LLM AS AN INTERPRETIVE MEDIATOR: A PARADIGM-BY-PARADIGM ANALYSIS

3.1. What Epistemological Status for the LLM?

Technically, an LLM is a statistical model trained on very large text corpora to predict the most probable continuation of a sequence of words. Four properties of this mode of operation have direct epistemological implications:

- **Stochasticity:** for the same prompt, the model may produce different responses, depending on the sampling parameters used.
- **Opacity:** the path leading from a piece of data to an interpretation cannot be observed, even when the model provides an apparent justification for its response.
- **Dependence on the training corpus:** the learned regularities reflect dominant texts, which are often English-language and Western. Less documented social realities, such as those expressed in Darija or Amazigh, are at risk of being poorly rendered.
- **Temporal instability:** models change from one version to the next, sometimes without users being informed, which compromises the later reproduction of an analysis.

These properties rule out treating the LLM as a simple neutral tool, like qualitative data analysis software that organises data without interpreting them. The LLM, by contrast, proposes interpretations. Conversely, it cannot be regarded as a "co-researcher" either: it has no intention to know, no experience of the field and no responsibility for the results.

We propose to describe it as an **interpretive mediator**. It stands between the data and the researcher and transforms the former into propositions of meaning according to learned regularities. These regularities function as an implicit theory that neither the researcher nor the model can make fully explicit. The epistemological question then becomes: how can a given paradigm integrate a mediation whose logic partly escapes it?

3.2. In the Positivist Paradigm: Reliability under Strain

Positivism values standardisation and replicability. At first sight, the LLM seems to be the natural ally of this paradigm. It applies a coding scheme to thousands of documents with a consistency that no human team can achieve, and it reduces certain biases linked to coder fatigue or mood.

This advantage, however, is conditional. Stochasticity and version instability directly threaten reliability: two runs of the same analysis may diverge, and an analysis carried out today may be impossible to reproduce a year from now. The claimed objectivity is also fragile, since the biases of the training corpus are built into every coding decision. The risk is what Messeri and Crockett [6] describe as an illusion of objectivity: taking as neutral a judgement that is not.

The case of "silicon samples" is even more delicate. Argyle et al. [2] show that some models convincingly reproduce the response distributions of demographic subgroups. This fidelity, however, remains conditional and must be demonstrated for each population. In a context poorly represented in the training data, such as rural banking customers in Morocco, the external validity of such a sample is highly doubtful. The LLM then reproduces a discursive average rather than a real population.

3.3. In the Critical Realist Paradigm: From Detecting Regularities to Explanation

For critical realism, observable regularities constitute only the surface of reality. The LLM excels precisely at detecting co-occurrences and patterns in the domain of the empirical. The danger lies in confusing these linguistic regularities with generative mechanisms. The fact that a model frequently associates "trust" and

"branch proximity" in a corpus of interviews says nothing, in itself, about the mechanism linking these two phenomena.

The LLM can nevertheless play a useful role in the abductive phase of research. It can suggest candidate mechanisms that the researcher might not have considered, provided these suggestions are treated as conjectures to be tested rather than as findings. Retroduction, which requires reasoning about the conditions of possibility of a phenomenon, remains theoretical work that the researcher cannot delegate.

3.4. In the Interpretivist Paradigm: The Risk of Flattening Meaning

Interpretivism seeks to understand situated meanings. It is probably the most exposed paradigm. The LLM has no access to lived context: it perceives neither a respondent's irony, nor the weight of a silence, nor the cultural charge of a local expression. It tends to produce "average" readings aligned with the most frequent meanings in its corpus. In doing so, it risks erasing precisely what interpretive research seeks to capture: the singularity of a point of view.

Credibility, the central criterion of this paradigm, relies on the researcher's prolonged engagement in the field and on the validation of interpretations by participants [17]. Neither of these operations can be entrusted to a model. The LLM can, however, serve as an **adversarial reader**: it can be asked to propose interpretations that differ from the researcher's, in order to test their robustness. It then does not produce the interpretation; it challenges it.

3.5. In the Pragmatic Constructivist Paradigm: An Interlocutor Subject to Traceability

Pragmatic constructivism appears to be the most welcoming paradigm. If knowledge is valued for its viability rather than for its correspondence with an independent reality, the LLM can legitimately take part in its construction, as one interlocutor among others. This paradigm nevertheless requires the researcher to justify each inference explicitly and to account for the process of elaboration [9].

The risk here is more subtle: outsourcing to the machine the conceptualisation work that is precisely where knowledge is produced. A researcher who adopts a category proposed by an LLM without being able to explain why it is relevant constructs nothing; he or she merely records. The condition of legitimacy is therefore the complete traceability of exchanges and the explicit statement of the reasons why each proposal was accepted or rejected.

3.6. Three Cross-Cutting Issues

Beyond the specificities of each paradigm, three issues cut across all positions. The first is **epistemic dependence**: the researcher places in the machine a trust comparable to that placed in testimony, without having the usual means of assessing its source. The second is the **shifting of responsibility**: when an erroneous interpretation is published, responsibility cannot be shared with a system that bears none. The third is the **homogenisation of knowledge**: if many researchers rely on the same models, their analytical categories are likely to converge, to the detriment of the theoretical diversity the social sciences need.

4. METHODOLOGY: A PROTOCOL FOR COMPARING HUMAN AND LLM CODING

The preceding conceptual analysis calls for empirical testing. We propose a reproducible protocol for comparing human and algorithmic coding of the same qualitative corpus. The protocol is presented here in its design; its implementation and results will be the subject of a later publication.

4.1. Corpus and Field

The intended field is the Moroccan banking sector, chosen for two reasons. First, it is undergoing a rapid digital transformation that generates rich discourse among managers. Second, interviews in this sector mix French, Standard Arabic and Darija, which makes it possible to test the models' sensitivity to local

linguistic realities. The planned corpus comprises some twenty to thirty semi-structured interviews with branch managers and regional executives about their experience of the digitalisation of customer relationships.

4.2. Comparison Design

The design is based on three coding conditions applied to the same corpus.

Table 2. Coding Conditions of the Comparison Protocol

Condition	Coder	Coding mode	Objective
A	Two researchers, working independently	Deductive thematic scheme, then emergent codes	Human benchmark and measurement of inter-coder agreement
B	LLM guided by the same scheme	Deductive, standardised prompt	Test the model's reliability against an explicit scheme
C	LLM with no prior scheme	Inductive, open prompt	Test the model's ability to generate emergent themes

Source: Author

Human coding follows the thematic analysis approach of Braun and Clarke [18]. The researchers in condition A code without knowing the results of conditions B and C, in order to avoid any contamination.

4.3. Indicators

Four families of indicators are planned:

1. **Inter-coder agreement**, measured by Krippendorff's alpha [19], between the two human coders, then between each human coder and the model.
2. **Intra-model stability**, measured by running the same prompt five times on the same corpus and comparing the results.
3. **Thematic coverage**, which compares the themes identified by the humans and by the model in the inductive condition.
4. **A typology of divergences**, built through qualitative analysis of the segments coded differently. Four categories are anticipated: contextual error, over-generalisation, invented theme with no grounding in the data, and relevant contribution not identified by the humans.

The fourth family of indicators is the richest from an epistemological standpoint. It makes it possible to move from a question of performance (does the model code "well"?) to a question of nature (how does its interpretation differ from that of a situated researcher?).

4.4. Reproducibility and Ethical Conditions

To ensure reproducibility, the protocol requires recording the exact version of the model and the date of use, the sampling parameters, and all prompts used, which will be published in an appendix. On the ethical side, interviews will be anonymised before being submitted to any model. Data processing must comply with Law No. 09-08 on the protection of individuals with regard to the processing of personal data. Participants' consent forms will explicitly mention the use of an AI tool in the analysis.

5. PROPOSAL: A RIGOUR FRAMEWORK FOR AI-ASSISTED RESEARCH

The analysis in Sections 3 and 4 leads to the proposal of a five-dimension rigour framework. This framework does not replace the criteria specific to each paradigm; it complements them when an LLM is involved in the research process.

5.1. The Five Dimensions

Traceability requires that every intervention of the model be documented: at which stage, with which prompt, for which result, and what decision the researcher drew from it.

Reproducibility requires that technical conditions be recorded (model version, date, parameters) and that the stability of results be checked through repeated runs.

Reflexivity requires the researcher to make explicit how the use of the model may have shaped his or her categories, questions or interpretations.

Adversarial validation requires that the model's outputs be confronted with independent human judgement, with other sources or with other models.

Interpretive accountability requires that every published interpretation be owned by a researcher able to justify its content without invoking the authority of the machine.

5.2. Specification by Paradigm

The five dimensions do not carry the same weight in every paradigm. Table 3 specifies, for each paradigm, the use of the LLM considered legitimate and the priority safeguards.

Table 3. Rigour Framework for LLM-Assisted Research, by Paradigm

Paradigm	Legitimate use of the LLM	Use to be avoided	Priority safeguards
Positivism / post-positivism	Large-scale deductive coding, pre-testing of instruments	Substituting simulated respondents for a real, undocumented population	Reproducibility, measured agreement with human coders
Critical realism	Generating candidate mechanisms in the abductive phase	Treating a co-occurrence detected by the model as a mechanism	Adversarial validation, interpretive accountability
Interpretivism	Adversarial reading of the researcher's interpretations	Delegating the interpretation of situated meanings	Reflexivity, validation by participants
Pragmatic constructivism	Interlocutor in conceptual elaboration	Adopting a category without being able to justify its relevance	Complete traceability, justification of inferences

Source: Author

5.3. A Standardised Statement of Use

To make the framework operational, we propose that all LLM-assisted research include a statement of use, integrated into the methodology section or placed in an appendix. It would specify six elements:

1. the epistemological paradigm claimed;
2. the stages of the research at which the model intervened;

3. the model used, its version, the date and the parameters;
 4. the full prompts;
 5. the validation procedures implemented;
 6. the researcher's decisions that followed or contradicted the model's suggestions.
- This statement gives reviewers a simple means of checking the coherence between the paradigm claimed and the actual use of AI.

6. DISCUSSION

6.1. Theoretical Contributions

The first contribution of the paper is to shift the debate on AI in research from the terrain of tooling to that of epistemology. The literature oscillates between two positions: enthusiasm, which sees LLMs as a gain in productivity and objectivity [1], and distrust, which sees them as a threat to scientific understanding [5, 6]. The paper shows that both positions are incomplete. The legitimacy of a given use depends on the paradigm within which it takes place. The same use may be acceptable in one position and problematic in another.

The second contribution is conceptual. The notion of an interpretive mediator avoids two symmetrical pitfalls: reducing the LLM to a transparent instrument and elevating it to the rank of a knowing subject. It extends the work on epistemological positioning in management science [8, 9] by incorporating a non-human actor that takes part in the production of meaning.

6.2. Practical Implications

For **researchers**, the proposed framework offers a decision frame: before using an LLM, one should ask whether this use is compatible with the paradigm claimed and which safeguards it requires. For **doctoral schools**, the paper argues for integrating the reasoned use of AI into methodology training, not as a technical skill but as an epistemological question. For **journals and reviewers**, the standardised statement of use provides a concrete evaluation tool.

The Moroccan and, more broadly, African context calls for particular vigilance. Dominant models are trained on corpora in which local languages and social realities are under-represented. The risk is that their widespread use invisibly imports analytical categories forged elsewhere. The question of research data sovereignty also arises when interviews are submitted to servers located outside the national territory.

6.3. Limitations and Avenues for Future Research

This paper has three limitations. The first stems from its conceptual nature: the propositions put forward must be tested empirically, which the protocol in Section 4 will make possible. The second relates to the speed at which models evolve: some of the properties described, such as version instability, could be mitigated by open and archivable models. The third concerns the scope adopted: the paper focuses on four paradigms and leaves aside critical and feminist approaches, which would shed specific light on the power relations embedded in the models.

Three avenues for future research emerge: implementing the comparison protocol on multilingual corpora; studying doctoral students' actual use of AI and its coherence with their declared paradigm; and analysing the effect of theoretical homogenisation at the level of a discipline.

7. CONCLUSION

The arrival of large language models in research practice is not a mere technical development. It forces the social sciences to return to their founding questions: what is valid knowledge, how is it produced, and

who answers for it? This paper has argued that the LLM should be understood as an interpretive mediator whose effects on rigour vary according to the paradigm adopted. Positivism gains processing capacity but loses demonstrable reliability. Critical realism finds support for abduction but must guard against confusing regularities with mechanisms. Interpretivism faces a risk of flattening meaning. Pragmatic constructivism can welcome the LLM as an interlocutor, provided that traceability is complete.

The five-dimension rigour framework and the statement of use proposed here aim to turn these findings into practice. Their common principle is simple: AI can assist the production of knowledge, but it cannot bear responsibility for it. Epistemological positioning, often perceived as an academic formality, thus becomes once again what it should always be: the compass that guides the researcher's methodological choices.

REFERENCES

1. F. Gilardi, M. Alizadeh, and M. Kubli, "ChatGPT outperforms crowd workers for text-annotation tasks," *Proceedings of the National Academy of Sciences*, vol. 120, no. 30, e2305016120, 2023.
2. L. P. Argyle, E. C. Busby, N. Fulda, J. R. Gubler, C. Rytting, and D. Wingate, "Out of one, many: Using language models to simulate human samples," *Political Analysis*, vol. 31, no. 3, pp. 337–351, 2023.
3. D. L. Morgan, "Exploring the use of artificial intelligence for qualitative data analysis: The case of ChatGPT," *International Journal of Qualitative Methods*, vol. 22, 2023.
4. C. Ziems, W. Held, O. Shaikh, J. Chen, Z. Zhang, and D. Yang, "Can large language models transform computational social science?," *Computational Linguistics*, vol. 50, no. 1, pp. 237–291, 2024.
5. E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the dangers of stochastic parrots: Can language models be too big?," in *Proc. ACM Conference on Fairness, Accountability, and Transparency (FAccT)*, 2021, pp. 610–623.
6. L. Messeri and M. J. Crockett, "Artificial intelligence and illusions of understanding in scientific research," *Nature*, vol. 627, no. 8002, pp. 49–58, 2024.
7. J. Piaget (Ed.), *Logique et connaissance scientifique*. Paris: Gallimard, 1967.
8. F. Allard-Poesi and V. Perret, "Fondements épistémologiques de la recherche," in R.-A. Thiétart (Ed.), *Méthodes de recherche en management*, 4th ed. Paris: Dunod, 2014.
9. M.-J. Avenier and C. Thomas, "Finding one's way around various methodological guidelines for doing rigorous case studies: A comparison of four epistemological frameworks," *Systèmes d'Information et Management*, vol. 20, no. 1, pp. 61–98, 2015.
10. J.-L. Le Moigne, *Les épistémologies constructivistes*. Paris: Presses Universitaires de France, 1995.
11. G. Burrell and G. Morgan, *Sociological Paradigms and Organisational Analysis*. London: Heinemann, 1979.
12. M.-J. Avenier, "Les paradigmes épistémologiques constructivistes : post-modernisme ou pragmatisme ?," *Management & Avenir*, no. 43, pp. 372–391, 2011.
13. K. R. Popper, *The Logic of Scientific Discovery*. London: Hutchinson, 1959.
14. R. Bhaskar, *A Realist Theory of Science*. Leeds: Leeds Books, 1975.
15. A. Sayer, *Method in Social Science: A Realist Approach*, 2nd ed. London: Routledge, 1992.
16. E. G. Guba and Y. S. Lincoln, "Competing paradigms in qualitative research," in N. K. Denzin and Y. S. Lincoln (Eds.), *Handbook of Qualitative Research*. Thousand Oaks, CA: Sage, 1994, pp. 105–117.
17. Y. S. Lincoln and E. G. Guba, *Naturalistic Inquiry*. Beverly Hills, CA: Sage, 1985.

18. V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006.
19. K. Krippendorff, *Content Analysis: An Introduction to Its Methodology*, 2nd ed. Thousand Oaks, CA: Sage, 2004.