A Framework For Discussing LLMs as Tools for Qualitative Analysis

James Eschrich
jamesae2@illinios.edu
University of Illinios, Urbana-Champaign
Urbana, Illinios, USA

Sarah Sterman
ssterman@illinois.edu
University of Illinios, Urbana-Champaign
Urbana, Illinios, USA

ABSTRACT

We review discourses about the philosophy of science in qualitative research and evidence from cognitive linguistics in order to ground a framework for discussing the use of Large Language Models (LLMs) to support the qualitative analysis process. This framework involves asking two key questions: "is the LLM **proposing** or **refuting** a qualitative model?" and "is the human researcher **checking** the LLM's decision-making directly?". We then discuss an implication of this framework: that using LLMs to *surface counter-examples for human review* represents a promising space for the adoption of LLMs into the qualitative research process. This space is promising because it is a site of overlap between researchers working from a variety of philosophical assumptions, enabling productive cross-paradigm collaboration on tools and practices.

CCS CONCEPTS

• **Human-centered computing** → **HCI design and evaluation methods**; *Collaborative and social computing design and evaluation methods*; *Field studies.*

KEYWORDS

Epistemology, Constructivism, Positivism, Post-positivism, Large Language Models, LLMs, LLM, Large Language Model, Qualitative Analysis, Qualitative Research, Grounded Theory, Cognitive Linguistics, Embodied Cognition

1 INTRODUCTION

LLMs can perform better than traditional NLP techniques at certain "natural language understanding" tasks like "miscellaneous text classification" [19], which raises the possibility that they could potentially be used for qualitative data analysis. Emergent literature investigates some possibilities such as using LLMs for deductive coding [17, 18] and inductive thematic analysis [5]. However, there is disagreement about whether, where, and to what extent LLMs

should or could be used to support the qualitative analysis process [8, 16].

In order to structure discussion about whether and under what conditions LLMs could be used to support the qualitative analysis process, we review relevant background on the philosophy of science, discussing **positivism** and **constructivism** using [14]'s language of *consistency* and *congruence*. We also review evidence from cognitive linguistics about how humans understand language. Then, based on this background, we propose a framework for structuring discussions around the use of LLMs to support qualitative analysis. This framework involves asking "is the LLM **proposing** or **refuting** a qualitative model?" and "is the human researcher **checking** the LLM's decision making directly?" Finally, we will discuss an implication of this framework: that *using LLMs to surface counter-examples for human review* from a qualitative dataset represents a promising space for the adoption of LLMs into the qualitative analysis process.

2 BACKGROUND

We characterize two positions on the philosophy of science, **positivism** and **constructivism**, using [14]'s language of *consistency* and *congruence*. We then review evidence from cognitive linguistics about how humans process language, and discuss the implications for using LLMs in qualitative analysis from both **positivist** and **constructivist** perspectives.

2.1 Positivism and Constructivism

Discourse about the philosophy of science in qualitative research has a long and complex history [1, 6]. A variety of "ways of knowing" [15] are taken up by researchers, who are often conceptualized as operating within "paradigms"—multiple perspectives on foundational philosophical issues packaged into coherent systems [12]. This picture is over-simplified [7] and potentially limiting [2], but we will further limit our discussion to two paradigms—**positivism** and **constructivism**—in reference to which many other paradigms, like **post-positivism** or [14]'s "constrained constructivism" are defined. We characterize **positivism** and **constructivism** using [14]'s language of *congruence* and *consistency*, which we find more useful here than [3, 12]'s language of ontology, epistemology, and methodology [1].

Positivism is oriented towards identifying and validating *congruent* models—models for which a "one-to-one correspondence" [14] can be established between the structure of the model and the structure of external reality. Since models which are *inconsistent*

---

[1]This is in part because some **constructivist** perspectives reject "the ontology/epistemology distinction" [12].

with observation cannot be *congruent*, discriminating between *consistent* and *inconsistent* models is relevant. However, neither the question of how to discriminate between different *consistent* models or the possible existence of *unknown* or unknowable models are of particular concern. This is because if *congruence* with reality is established, the possibility spaces of the *consistent* and the *unknown* become irrelevant.

Constructivism denies the possibility of establishing *congruent* models, instead radically acknowledging induction's inherent limits [12] and the *unknown*—models not yet tested, not yet conceived, or fundamentally unknowable by humans [14]. **Constructivism** thus reverses the emphases of **positivism**, focusing on exploring the possibility space of *consistent* models while engaging reflexively with the limits of our perspectives and, thus, the omnipresent *unknown*.

The language of *consistency* and *congruence* helps reveal these paradigms as two points on a continuum; different positions can be located by examining how much attention is paid to *congruent* models over *unknown* models or *consistent* models over *inconsistent* models. This language also highlights an important area of overlap between **constructivism** and **positivism**. Both paradigms are interested in rejecting models that are *inconsistent* with observation: the **positivist**[2] while trying to establishing *congruence* with reality and the **constructivist** while balancing an exploration of different *consistent* models with a continual deference to the *unknown*.

2.2 LLMs and The Implications of Cognitive Linguistics

Evidence increasingly suggests that the human ability to understand natural language is based on both cultural and embodied experience [10, 13]. This poses a strong challenge to any attempts to use LLMs in the qualitative analysis process. However, the nature of this challenge varies across philosophical frameworks. For **positivists**, concerned with discovering and validating models congruent with reality, the process by which LLMs produce claims is not intrinsically important; the structure of the reality we are trying to model exists independently and can theoretically be arrived at through multiple means. However, attempting to measure how consistently LLMs arrive at congruence with reality remains important.

For **constructivists**, concerned with discriminating between consistent models and keeping the unknown in view, the process by which LLMs make claims about language data is our primary concern. Constructivists understand knowledge as enabled by "particular sets of sensory apparatus located within specific cultures and times" [14]; thus, the constructivist's ability to evaluate claims produced by an LLM requires understanding how the LLM came to produce those claims. What value, if any, claims originating from LLMs have from this perspective is uncertain and unclear.

3 A FRAMEWORK FOR DISCUSSING LLMS FOR QUALITATIVE ANALYSIS

With the philosophical and linguistic background reviewed, we propose a framework (Table 1) for discussing the use of LLMs for qualitative analysis. This framework is structured around two questions. The first question, "is the LLM **proposing** or **refuting** a qualitative model?" highlights the important "asymmetry" [14] between affirming consistency and refuting consistency. The second question, "is the human researcher **checking** the LLM's decision-making directly?" draws attention to the complex processes by which humans come to understand language and to what extent LLMs are capable of either supporting or replacing those processes.

4 DISCUSSION

The upper-right quadrant of Table 1, which we will call 'Using LLMs to *surface counter-examples for human review*' is unique in that, in this case, the positivists and constructivists have the same questions. This is because both positivists and constructivists are interested in discarding models which are inconsistent with observation, although they may differ on what sort of evidence implies inconsistency and in what contexts that inconsistency holds.

We can also articulate this point using the language of ontology, epistemology, and methodology. Negating a model is ontologically neutral; 'how can we judge whether this model is inconsistent with observation?' is a purely epistemological question. Moreover, when a human researcher is the one answering that question, using an LLM to support the process is not *inherently* problematic epistemologically; no tool for performing qualitative analysis is neutral, and all tools influence the process in some way [11]. Thus, only the methodological question remains: is using LLMs to *surface counter-examples for human review* useful/effective for doing qualitative analysis?

One way using LLMs to *surface counter-examples for human review* could be useful is by enabling researchers to work with very large datasets. Existing mixed-methods work has used topic modeling to "purposively sample" from very large datasets like posts on a subreddit [9]. This is an exciting approach; however, topic modeling represents a situation in which a language model is '**proposing** a qualitative model'. As an alternate approach, an initial sample could be selected for a preliminary qualitative analysis using traditional methods like random or stratified sampling. This preliminary analysis could then be refined by using an LLM to *surface counter-examples for human review* from the much larger dataset. This somewhat resembles the use of theoretical sampling in grounded theory to "develop and refine" emerging categories [4, p. 205].

How using LLMs to *surface counter-examples for human review* might positively or negatively affect the results of a qualitative analysis is an open question. However, restricting the LLM to *surfacing counter-examples for human review* ensures that (a) all claims about which qualitative models are consistent with observation originate from human researchers and (b) no qualitative model is rejected by an LLM-identified observation without giving human researchers a chance to revise the model to account for it. This creates a point of

---

[2]In this paper, 'the positivist' and 'the constructivist' represent "ideal ... and extreme" [2, p. 5] characterizations of the research paradigms deployed for explanatory purposes, rather than individuals who may use or identify with a paradigm for specific reasons within specific contexts.

Table 1: A Framework for Discussing LLMs for Qualitaive Analysis

	LLM is **proposing** a qualitative model (*e.g.* codes, themes, categories, clusters)	LLM is **refuting** a qualitative model (*e.g.* counter-examples)
Human researcher is **checking** the LLM's decisions directly	*Example*: clustering related codes for human review, automated/suggested open coding with human review *Positivist*: How is this supporting or hindering my ability to validate congruence with reality? *Constructivist*: Do I agree the LLM's proposed model is consistent with these observations? How is this supporting or hindering my ability to explore other consistent models? How is this affecting my ability to reflect on my experience?	*Example*: suggesting counter-examples to a developing theory/category/theme *Positivist*: Should this model be rejected as inconsistent? Can I account for this counter-example? *Constructivist*: Should this model be rejected as inconsistent? Can I account for this counter-example?
Human researcher is **NOT checking** the LLM's decisions directly	*Example*: automated open coding without human review *Positivist*: How can I be confident that the LLM's claims are congruent with reality? If they aren't, are they similar enough to an "average human's" claims to be useful? *Constructivist*: How can we evaluate or use a qualitative model that does not arise from an inter-subjective human experience? Should we be doing this in the first place?	*Example*: deductive coding without human review (i.e. for each observation, eliminating codes which are inconsistent) *Positivist*: How well does an LLM perform compared to a human in correctly identifying data inconsistent with a given model? *Constructivist*: How do the inconsistencies an LLM identifies compare with mine? Other people's? Is the LLM biased in a certain way? How is this shaping my analysis? Should we be doing this in the first place?

compatibility which could enable tools and practices to be collaboratively developed and productively employed across philosophical boundaries.

5 CONCLUSION

Based on a review of the philosophy of science in qualitative research and evidence from cognitive linguistics, we propose a framework for discussing the use of LLMs for qualitative analysis. This framework is based on two key questions: "is the LLM **proposing** or **refuting** a qualitative model?" and "is the human researcher **checking** the LLM's decision-making directly?". We discuss the potential of using LLMs to *surface counter-examples for human review*, which this framework identifies as promising because it is a site of overlap between different philosophical perspectives. We hope our framework and discussion will help researchers working from different research paradigms discuss the potential of LLMs as tools for supporting qualitative analysis.

REFERENCES

[1] Svend Brinkmann, Michael Hviid Jacobsen, and Søren Kristiansen. 2014. Historical Overview of Qualitative Research in the Social Sciences. In *The Oxford Handbook of Qualitative Research*, Patricia Leavy (Ed.). Oxford University Press, 0. https://doi.org/10.1093/oxfordhb/9780199811755.013.017

[2] Roger Chafe. 2023. Rejecting Choices: The Problematic Origins of Researcher-Defined Paradigms within Qualitative Research. *International Journal of Qualitative Methods* 22 (Jan. 2023), 160940692311659. https://doi.org/10.1177/16094069231165951

[3] Kerry Chamberlain. 2000. Methodolatry and Qualitative Health Research. *Journal of Health Psychology* 5, 3 (May 2000), 285–296. https://doi.org/10.1177/135910530000500306

[4] Kathy Charmaz. 2014. *Constructing Grounded Theory* (second edition ed.). SAGE Publications Ltd, London ; Thousand Oaks, Calif.

[5] Stefano De Paoli. 2023. Performing an Inductive Thematic Analysis of Semi-Structured Interviews With a Large Language Model: An Exploration and Provocation on the Limits of the Approach. *Social Science Computer Review* (Dec. 2023), 08944393231220483. https://doi.org/10.1177/08944393231220483

[6] Norman K. Denzin and Yvonna S. Lincoln. 2005. Introduction: The Discipline and Practice of Qualitative Research. In *The Sage Handbook of Qualitative Research, 3rd Ed.* Sage Publications Ltd, Thousand Oaks, CA, 1–32.

[7] Norman K. Denzin and Yvonna S. Lincoln. 2017. *Introduction: The Discipline and Practice of Qualitative Research* (5th ed.). SAGE Publications, Chapter 1.

[8] Jessica L. Feuston and Jed R. Brubaker. 2021. Putting Tools in Their Place: The Role of Time and Perspective in Human-AI Collaboration for Qualitative Analysis. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 469:1–469:25. https://doi.org/10.1145/3479856

[9] Robert P Gauthier, Mary Jean Costello, and James R Wallace. 2022. "I Will Not Drink With You Today": A Topic-Guided Thematic Analysis of Addiction Recovery on Reddit. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–17. https://doi.org/10.1145/3491102.3502076

[10] Raymond W. Gibbs, Paula Lenz Costa Lima, and Edson Francozo. 2004. Metaphor Is Grounded in Embodied Experience. *Journal of Pragmatics* 36, 7 (July 2004), 1189–1210. https://doi.org/10.1016/j.pragma.2003.10.009

[11] Linda S. Gilbert, Kristi Jackson, and Silvana di Gregorio. 2014. Tools for Analyzing Qualitative Data: The History and Relevance of Qualitative Data Analysis Software. In *Handbook of Research on Educational Communications and Technology*, J. Michael Spector, M. David Merrill, Jan Elen, and M. J. Bishop (Eds.). Springer, New York, NY, 221–236. https://doi.org/10.1007/978-1-4614-3185-5_18

[12] Egon G. Guba. 1990. The Alternative Paradigm Dialog. In *The Paradigm Dialog*. Sage Publications, Inc, Thousand Oaks, CA, US, 17–27.

[13] Beate Hampe. 2017. Embodiment and Discourse: Dimensions and Dynamics of Contemporary Metaphor Theory. In *Metaphor: Embodied Cognition and Discourse*, Beate Hampe (Ed.). Cambridge University Press, Cambridge, 3–24. https://doi.org/10.1017/9781108182324.002

[14] N. Katherine Hayles. 1993/1993. Constrained Constructivism: Locating Scientific Inquiry in the Theater of Representation. In *Realism and Representation: Essays on the Problem of Realism in Relation to Science, Literature, and Culture*, George Levine (Ed.). University of Wisconsin Press, Madison, 27–43.

[15] Wendy A. Kellogg and Judith S. Olson. 2014. Epilogue. In *Ways of Knowing in HCI*, Judith S. Olson and Wendy A. Kellogg (Eds.). Springer, New York, NY, 469–472. https://doi.org/10.1007/978-1-4939-0378-8_19

[16] John Roberts, Max Baker, and Jane Andrew. 2024. Artificial Intelligence and Qualitative Research: The Promise and Perils of Large Language Model (LLM) 'Assistance'. *Critical Perspectives on Accounting* 99 (March 2024), 102722. https://doi.org/10.1016/j.cpa.2024.102722

[17] Robert H. Tai, Lillian R. Bentley, Xin Xia, Jason M. Sitt, Sarah C. Fankhauser, Ana M. Chicas-Mosier, and Barnas G. Monteith. 2024. An Examination of the Use of Large Language Models to Aid Analysis of Textual Data. *International Journal of Qualitative Methods* 23 (Jan. 2024), 16094069241231168. https://doi.org/10.1177/16094069241231168

[18] Ziang Xiao, Xingdi Yuan, Q. Vera Liao, Rania Abdelghani, and Pierre-Yves Oudeyer. 2023. Supporting Qualitative Analysis with Large Language Models: Combining Codebook with GPT-3 for Deductive Coding. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 75–78. https://doi.org/10.1145/3581754.3584136

[19] Jingfeng Yang, Hongye Jin, Ruixiang Tang, Xiaotian Han, Qizhang Feng, Haoming Jiang, Shaochen Zhong, Bing Yin, and Xia Hu. 2024. Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond. *ACM Transactions on Knowledge Discovery from Data* 18 (Feb. 2024), 1–32. Issue 6. https://doi.org/10.1145/3649506