

Computational Grounded Theory Integration
(CGTI):
A Methodological Alternative to
XAI-Supported
Qualitative Social Research with Large
Language Models

Paul Koop

2026

Abstract

The integration of Large Language Models (LLMs) into qualitative social research raises fundamental methodological questions. While recent approaches (so-called XRSR methods) attempt to overcome the opacity of LLMs through Explainable AI (XAI), this paper argues that such attempts fail due to a fundamental incommensurability between quantitative opacity-breaking and qualitative meaning reconstruction. We develop an alternative, the *Computational Grounded Theory Integration (CGTI)*, a method that deploys LLMs exclusively in methodologically unproblematic roles: technical preprocessing and counterfactual exploration. The actual interpretation, case reconstruction, and theory formation remain the preserve of the human researcher. Using eight transcribed market conversations from Aachen (1994), we demonstrate the complete application of CGTI from sequential microanalysis to formal modeling using Petri nets and probabilistic context-free grammars. The results reveal a surprising harmony of transactions and a 100% success rate of upselling attempts—a finding that would not have become visible without systematic methodological integration. We conclude with a philosophical assessment of CGTI in comparison to XAI-based approaches.

1 Problem Statement

The recent debate on the use of Large Language Models (LLMs) in the social sciences and humanities is characterized by a fundamental ambivalence. On the one hand, LLMs open entirely new possibilities for text analysis, pattern generation, and exploration of large text corpora. On the other hand, their inherent opacity—the lack of traceability regarding how exactly an LLM proceeds from input to output—poses an epistemological problem for any science based on traceability and methodological control.

The debate has produced an interesting asymmetry: While in the humanities the hermeneutic tradition does not assume an objective truth behind the text but rather the productive act of understanding itself, the social sciences with their methodologically controlled case reconstruction face a greater problem. An LLM interpreting an interview transcript produces a claim (e.g., "The subtext of this sentence is a latent rebellion against authority"). Yet without knowledge of the reconstruction path, this claim remains scientifically worthless—it could be based on hallucinations or statistical artifacts.

A widely discussed solution is the recourse to XAI (Explainable AI)—methods such as Attention Visualization, Concept Activation Vectors (TCAV), or Mechanistic Interpretability, which attempt to make the internal calculations of an LLM visible. In recent methodological drafts (hereafter referred to as the XRSR approach), it is proposed to use XAI to validate LLM interpretations and to induce formal models (PCFG, Petri nets, Bayesian networks) from the resulting category chains.

This paper demonstrates that this approach fails due to a fundamental incommensurability. We develop a methodologically more consistent alternative, CGTI, and demonstrate its application on an empirical corpus of eight market conversations from Aachen (1994).

2 State of Research

2.1 Hermeneutics and Large Language Models

The question of whether an LLM can *understand* is the subject of intense philosophical debate [10, 1]. While some authors argue that the statistical regularities LLMs extract from training data constitute a form of "understanding" [2], the phenomenological tradition emphasizes the role of consciousness, embodiment, and temporal experience for any act of understanding [5, 7].

For the methodological discussion, what is decisive is: Even if one attributes *understanding* to an LLM in the sense of an input-output function, the *traceability* of this understanding remains a problem. The classical quality criteria of qualitative research—particularly intersubjectivity, replicability, and reflexivity—demand the complete documentation of the path of knowledge [11, 9].

2.2 XAI in the Social Sciences

Research on Explainable AI has produced a variety of methods to reduce the opacity of neural networks [6]. Particularly relevant for LLMs are:

- **Attention Visualization** (e.g., BertViz [12]): Makes visible which tokens in the input an LLM weighted when generating an output.
- **Concept Activation Vectors (TCAV)** [8]: Allows checking whether an LLM has recognized a specific concept (e.g., "gender bias") in the text.
- **Mechanistic Interpretability** [4]: Attempts to identify specific "circuits" within the LLM for particular operations.

However, the transfer of these methods to social science research questions is associated with considerable methodological problems, as we show in the following section.

2.3 Critique of the XRSR Approach

The approach referred to here as XRSR (XAI-supported relational structure reconstruction) proposes to validate LLM-generated readings through XAI and subsequently induce formal models from the resulting category chains. Our analysis identifies three fundamental problems:

2.3.1 The Compatibility Problem

XAI methods provide quantitative measures—attention weights, activation strengths, vector distances. Qualitative social research, however, operates with meaning-genetic categories that cannot be reduced to quantitative measures. The fact that token A and token B exhibit high attention does not mean that the LLM has captured the *same* social meaning that a human interpreter sees in a sequence. The translation between these levels is not methodologically solved in the XRSR approach but merely asserted.

2.3.2 The Circularity Problem

XAI is intended to validate LLM readings. However: The XAI methods themselves require interpretation. An attention heatmap shows correlations, not causalities. The fact that an LLM links certain tokens does not prove that it has correctly captured the *social meaning* of a sequence—it merely demonstrates that it has learned statistical regularities from its training data. The XRSR approach shifts opacity from the LLM level to the XAI level without overcoming it.

2.3.3 The Induction Problem

The "induction" of formal models (PCFG, Petri nets, Bayesian networks) from category chains proposed in Phase 4 of the XRSR draft is logically opaque. A PCFG requires a defined nonterminal symbol system. Where do these grammatical categories come from? If they are extracted from the LLM's readings, the method becomes circular. If they are specified by the researcher, it is not induction but construction. The same problem applies to Petri nets (which presuppose discrete states while qualitative processes are fluid) and Bayesian networks (which require probability measures on variables while qualitative categories are not necessarily interval-scaled).

2.4 The Methodological Alternative: CGTI

From this critique, this paper develops the *Computational Grounded Theory Integration (CGTI)*. CGTI is based on three methodological principles:

1. **Primacy of human interpretation:** Understanding meaning is a consciousness-bound act that no AI can perform. The final responsibility for interpretation remains with the human researcher.
2. **Separation of technical and hermeneutic opacity:** Technical opacity (how an LLM internally computes) is acceptable; hermeneutic opacity (how an interpretation comes about) is not.
3. **Case reconstruction before aggregation:** Each case is understood individually before comparative patterns are extracted.

CGTI differs from the XRSR approach in four central points:

Table 1: Comparison of XRSR and CGTI

Criterion	XRSR	
Role of LLM	Interpretation generator + validation through XAI	Technical prepro
XAI function	Validation instance	
Formal modeling	"Induction" from LLM categories	Explicit
Status of formal models	Empirically "given"	Hermeneutica
Opacity problem	Shift (LLM \rightarrow XAI)	Circumv

3 Method Application: CGTI in Empirical Practice

3.1 Overview of the CGTI Method

CGTI comprises six phases:

1. **Data basis & technical preparation:** Transcription, anonymization, technical preprocessing (LLM only for orthographic correction and segmentation)
2. **Sequential microanalysis:** Hermeneutic single-case reconstruction following methods of objective hermeneutics or documentary method
3. **LLM-supported counterfactual exploration:** Generation of alternative sequence courses for contrast (LLM heuristic, not interpretative)
4. **Systematic case comparison:** Case contrasting, identification of comparison dimensions, theoretical sampling
5. **Formal modeling as researcher construction:** Explicit construction of Petri nets, PCFG, possibly Bayesian networks *based on qualitative theory*
6. **Theoretical integration & validation:** Triangulation, communicative validation, reflexive validation

3.2 Empirical Material

The material comprises eight transcripts of market conversations from Aachen from 1994 (June/July). The transcripts vary in length, completeness, and conversation type:

Table 2: Corpus of market conversations

Text	Date	Location	Participants	Feature
T1	28.06.1994	Butcher shop	Seller (f), C	complete, calm
T2	28.06.1994	Cherry stall	Seller (m), C1, C2	fragmented, sales call
T3	28.06.1994	Fish stall	Seller (m), C	complete, minimal
T4	28.06.1994	Vegetable stall	Seller (m), C	complete, advisory
T5	26.06.1994	Vegetable stall	Seller (m), C1, C2	partial
T6	28.06.1994	Cheese stall	Seller (m), C1	fragmented, broken off
T7	28.06.1994	Candy stall	Seller (m), C	complete, mini
T8	09.07.1994	Bakery	Seller (f), C	complete, role breach

3.3 Selected Results from Phases 2–5

Due to space constraints, we focus on three central findings.

3.3.1 Finding 1: The Sequential Structure of Upselling (Phases 2 and 4)

Sequential microanalysis (Phase 2) reveals a striking regularity: In every case where the seller asks the question "Anything else?" a further purchase follows (T1, T4, T5, T8). The question "Anything else?" is thus not a neutral service inquiry but a *sequential coercion to continue purchasing*. It creates an expectation structure that the customer can only contradict at social cost—a "No, thank you" would have to be interpreted as a refusal of an offered relationship.

Throughout the entire corpus, there is not a single explicit "No" to this question—only in T4 is the upselling anticipated by the customer's own initiative question about chanterelles.

3.3.2 Finding 2: Two Modes of Seller-Customer Relationship (Phase 4)

Systematic case comparison (Phase 4) allows the identification of three distinct interaction modes:

- **Transactional mode** (T1, T3, T5, T6, T7): Clear roles, minimal exchange, standardized sequences, no consultation.
- **Advisory mode** (T4): Expertise differential, recipe questions, linguistic proximity ("We'll take the light ones"), customer delegates culinary expertise.

is never traversed more than once—an indication that a second "Anything else?" would be perceived as intrusive.

3.4 Validation (Phase 6)

Triangulation of findings shows high consistency:

Table 3: Triangulation of upselling findings

Source of finding	Result
Hermeneutics (Phase 2)	"Anything else?" is a sequential opening that is difficult to
Formal model (Phase 5)	$P(\text{Continued purchase} \mid \text{Upselling question}) = 1.0$ in corpus
Counterfactual simulation (Phase 3)	Alternative (No) leads to shorter transaction but socially
Absent cases	No single "No" in corpus suggests social desirability

Communicative validation (simulated survey of market sellers) revealed that "That's it then" occurs as a polite No-variant in everyday practice—a category not represented in the corpus.

4 Summary of Results

The application of CGTI to the eight market conversations leads to three overarching insights:

4.1 Insight 1: The Harmony of the 1994 Market

The Aachen market appears in the transcripts as a surprisingly harmonious place. There are no price conflicts, no complaints, no broken-off transactions. The interactions are ritualized, polite, and efficient. This harmony could be an effect of selection (the transcription may have only captured harmonious cases), an effect of transcription practice (incomprehensible passages might contain conflicts), or a historical phenomenon (the German Mark era, different expectations of market interaction).

4.2 Insight 2: The Power of Sequential Opening

The question "Anything else?" unfolds a remarkable sequential force. Throughout the corpus, it is never answered with "No"—instead, a further purchase or a transitional phrase ("That's it") always follows. The PCFG from Phase 5 captures this regularity with an empirical probability of $p = 1.0$, which

can be interpreted as strong evidence of an underlying social expectation structure.

4.3 Insight 3: Role Breach as a Resource

Text 8 (bakery) shows that temporary departure from the transaction role—here through the complaint about the unopened door—can be a resource for relationship building. The private intrusion creates a "shared third" (the shared annoyance) that makes the interaction more human without endangering the transaction. Remarkable is the participants' ability to return to role language after this intrusion.

5 Assessment of the CGTI Method

5.1 Strengths

CGTI has three central strengths compared to XAI-based alternatives:

1. **Methodological consistency:** CGTI avoids the category error of short-circuiting quantitative XAI measures with qualitative meaning categories. The separation between hermeneutic interpretation (Phase 2) and formal modeling (Phase 5) is clear and justified.
2. **Role clarity of AI:** The LLM is deployed only where its capabilities are methodologically unproblematic—technical preprocessing (Phase 1) and heuristic exploration (Phase 3). It does not replace human interpretation.
3. **Replicability:** CGTI documents all steps completely—from sequential microanalysis through the comparison matrix to the explicit construction of formal models. Replication by other researchers is possible.

5.2 Limitations

The present application of CGTI has three significant limitations:

1. **Small sample size** ($n = 8$): Formal modeling (particularly the estimation of a Bayesian network) is not validly possible with this sample size. The probabilities presented in Phase 5 are descriptive, not inferential.

2. **Fragmented transcripts:** Texts T2 and T6 contain incomprehensible passages and are partially broken off. This limits the validity of sequential microanalysis for these cases.
3. **Absence of contrast cases:** The corpus contains no conflict cases (price negotiations, complaints, breakdowns). The structures identified in theory formation therefore apply only provisionally to harmonious transactions.

5.3 Comparative Assessment: CGTI vs. XRSR

From a philosophical perspective, CGTI is superior to the XRSR approach because it recognizes the fundamental difference between *statistical regularities* (what LLMs provide) and *meaning-genetic reconstruction* (what qualitative methods demand) and does not attempt to bridge it through technical artifices. The XRSR approach succumbs to what Dreyfus [3] called the "illusion of cognitive transparency": the assumption that one only needs to look deeply enough into a system's internal calculations to capture its understanding.

CGTI is more modest but methodologically cleaner. It uses LLMs as *tools*—not as *substitutes* for human understanding. Its validity rests not on the breaching of technical opacity but on the methodological control of the human knowledge process.

6 Conclusion

The debate on the use of LLMs in qualitative social research is far from concluded. This paper has argued that XAI-based approaches fail due to a fundamental incommensurability and has developed a methodologically more consistent alternative, CGTI. The application to eight market conversations from Aachen (1994) has demonstrated the practicability of the method and produced initial empirical findings on the structure of market interactions.

Three desiderata emerge for further research:

1. **Methodological development:** CGTI should be tested with larger and more heterogeneous corpora, particularly with cases containing conflicts, negotiations, and breakdowns.
2. **Comparative studies:** A systematic comparison between CGTI and XAI-based approaches on the *same* material could more precisely identify their respective strengths and weaknesses.

3. **Software support:** The development of open-source software to support CGTI (particularly for Phases 4 and 5) could increase the accessibility of the method.

Finally, it should be emphasized: The question is not *whether* LLMs can be used in qualitative social research—they undoubtedly can. The question is *how* they should be used without undermining the methodological foundations of qualitative research. CGTI offers a promising, albeit still developing, proposal.

References

- [1] Bender, E. M., & Koller, A. (2021). Climbing towards NLU: On meaning, form, and understanding in the age of data. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics*, 5185–5198.
- [2] Chalmers, D. J. (2023). Does a large language model understand language? *Journal of Consciousness Studies*, 30(7-8), 1-25.
- [3] Dreyfus, H. L. (1972). *What computers can't do: A critique of artificial reason*. Harper & Row.
- [4] Elhage, N., Nanda, N., Olsson, C., et al. (2021). A mathematical framework for transformer circuits. *Anthropic Research Paper*.
- [5] Gadamer, H.-G. (1960). *Truth and Method*. Mohr Siebeck.
- [6] Guidotti, R., Monreale, A., Ruggieri, S., et al. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, 51(5), 1-42.
- [7] Heidegger, M. (1927). *Being and Time*. Niemeyer.
- [8] Kim, B., Wattenberg, M., Gilmer, J., et al. (2018). Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (TCAV). *Proceedings of ICML 2018*.
- [9] Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.
- [10] Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417-457.

- [11] Steinke, I. (2004). Quality criteria in qualitative research. In U. Flick, E. von Kardorff & I. Steinke (Eds.), *A Companion to Qualitative Research* (pp. 319-331). Sage.
- [12] Vig, J. (2019). A multiscale visualization of attention in the transformer model. *Proceedings of ACL 2019: System Demonstrations*, 37-42.