Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Machine-mediated theorizing: Examining epistemic work in a generative AI grounded theory workflow

**Melike Şahinol**

sahinol@oiist.org

Orient-Institut Istanbul

**Kocager Uğur**

Orient-Institut Istanbul

**Gülşah Başkavak**

[1] Orient-Institut Istanbul, “Science, Technology and Society”, Istanbul, Türkiye [2] Acıbadem University, Department of Sociology, Istanbul, Türkiye

**Ayşe Berna Uçarol**

Orient-Institut Istanbul

**Article**

**Keywords:** Epistemic Work, Grounded Theory, Sociotechnical Workflow, Generative AI, Qualitative Analysis

**Posted Date:** August 4th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-10526750/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# Abstract

Discussions on generative AI (GenAI) in knowledge-intensive fields continue to be shaped by polarized narratives that oscillate between promises and dangers. These narratives rarely address the context-dependent and historically contested epistemic practices of qualitative research, into which GenAI is entering. Grounded Theory (GT) provides a revealing case for examining this problem, since its establishment has been accompanied by a long-standing controversy over how theory emerges from empirical material and when methodological procedure turns into forcing.

This article examines how GenAI performs epistemic work when it is instructed not only to classify interview material but to transform it into codes, categories, relations, a core category, and a theoretical model. The study draws on three interviews from a research project on surgical instrument manufacturing, as they represent different organizational positions and forms of knowledge. An LLM was instructed to conduct a sequential workflow of open, axial, and selective coding without access to the research team's existing categories or interpretations.

The results generated by GenAI (e.g., codes, memos, paradigm models, etc.) were analyzed using a theory-guided qualitative content analysis. It became apparent that the model generated empirically verifiable conceptual condensations and relationally complex theoretical statements. It further favored linear, product-oriented process narratives, established coherence between scattered passages, and considered contradictory material more as a condition or limitation of scope rather than allowing it to fundamentally reorganize the emerging theory.

GenAI does not operate as a neutral coding tool nor as an autonomous theory generator. This is because its epistemic implications are embedded in a sociotechnical context in which, e.g., prompting, methodological rules, computational operations, and researcher evaluation are interwoven within a human-machine collaborative network. By reconstructing these processes, we can understand how machine-assisted abstraction redefines the distinction between emergence and forcing that has historically shaped GT.


# 1. Introduction

Existing research on large language models (LLMs) in qualitative analysis has expanded rapidly over the last few years. Two tendencies in this literature are particularly relevant to the problem addressed here. In the first, AI-generated and human-generated codings are compared in their accuracy and reliability with the conclusion that human intervention remains indispensable when segments of text are ambiguous, context-dependent, or theoretically complex (Hodges et al., 2025, Xiao et al., 2023, Han et al., 2025, Jayawardana et al., 2026, Bianco et al., 2026). The second examines the collaborative relationship between researchers and models, showing that the human role does not disappear but is redistributed in the analytical process, e.g. through iterative prompt design (Khalid and Witmer, 2025, Ramanathan et al., 2025, Sharma et al., 2025) or through the researcher's changing position as coder, mediator, and reflexive practitioner (Breazu et al., 2024, Park et al., 2025, Carmona-Díaz et al., 2026, Thominet et al., 2024).

Overall, the literature discussed presents qualitative coding as a technical procedure to be evaluated on its results, rather than as a historically contested epistemic practice whose methodological controversies could serve as a diagnostic tool for understanding what GenAI actually does when it theorizes.

These studies largely leave the epistemic practice in which these procedures are embedded in the background. Yet this is of crucial importance for GT, a method for organizing the relationships between empirical material, methodological rules, theoretical sensitivity, and the emerging subject of knowledge, and not only for assigning codes to empirical material. Since its formulation, GT has been accompanied by a controversy over how theory can be generated from empirical material, how much analytical guidance a researcher may use, and when such guidance begins to force material into categories that do not emerge from it (Glaser, 1992, Strauss and Corbin, 1990, Glaser and Strauss, 1967).

However, this issue was not resolved, as Charmaz's (2006) constructivist reformulation and Clarke's (2005) situational analysis shifted this controversy to questions of situational interpretation, power, silence, and the conditions under which categories acquire meaning. GT is therefore also a qualitative methodology that is a historically contested epistemic practice.

These controversies within GT raise the question of what epistemic work GenAI undertakes when it is expected not only to classify text passages but also to form categories, establish relationships between them, and develop theoretical memos. The question, therefore, is not whether an LLM can code like a human researcher, it is about what kind of theoretical abstraction emerges when the work of comparing, revising, and forming categories is partially delegated to a machine.

Debates about GenAI in knowledge-intensive fields continue to oscillate between promise and danger, and between technological and social explanations of what these systems do. Rather than adjudicating between these positions, we locate the question in a site where such binaries are rarely examined, namely in the epistemic practice of qualitative social research itself. In doing so, the article contributes to an emerging AI Social Science that treats machine-assisted analysis as co-produced within sociotechnical arrangements rather than as an autonomous technical capacity.

Therefore, this article asks how GenAI performs epistemic work within a GT workflow when it is instructed to theorize from empirical material, and what forms of theoretical abstraction this produces. More detailed, we examine how the model transforms empirical segments into concepts and categories, relates and revises categories, handles ambiguity and contradiction, and integrates categories into more general theoretical propositions.

## 2. Theoretical framework

# 2.1 Epistemic practice: How theory comes into being

Knowledge is never simply discovered. It is made and made differently depending on where and by whom. Knorr-Cetina's comparative ethnographies of high-energy physics and molecular biology in laboratories trace what she calls the "contemporary machineries of knowing" (Knorr-Cetina, 1999), showing that even within the natural sciences, there is no single, unified method by which knowledge is produced, but distinct epistemic cultures, each with its own instruments, evidentiary standards, and relationship between theory and empirical material (Knorr-Cetina, 1999). This variability holds even greater force for qualitative social research. In qualitative research, theory is not derived from data through a fixed procedure, but is built up gradually through repeated engagement with the empirical material, comparison across cases, and a sensitivity that must be cultivated over time.

Our use of the term epistemic work draws on several related traditions while distinguishing between them. The pragmatist sociology of work and action in the Straussian trajectory directs attention to the processual and coordinated activities through which heterogeneous tasks are articulated within a course of action (Strauss, 1993, Strauss, 1985, Strauss, 1988). Cook and Brown (1999) define directly the term epistemic work as the practical activity through which what is known is acquired, deployed, confirmed, or modified in the course of action. Yet both traditions principally formulate knowing in relation to human practice and as an aspect of action (or practice) that does epistemic work (Cook and Brown, 1999). Applying the concept to GenAI therefore requires neither treating the model as a human-like knower nor reducing it to a neutral instrument.

Given this background, we explore how such epistemic work is produced, enabled, and brought to bear within an analytical, sociotechnical context in which actors, materials, rules, and infrastructures interact during various phases of research (Strauss, 1988, Strauss, 1985, Orlikowski, 2007, Suchman, 2007). We understand epistemic work as a series of situational and materially mediated processes through which the human-machine division of labor selects, transforms, compares, relates, and examines empirical material, and provisionally establishes it as concepts, categories, and statements of knowledge (Suchman, 2011, Cook and Brown, 1999, Knorr-Cetina, 1999). The epistemic purpose of the research, the creation of the corpus and the prompts, the assessment of the appropriateness of emerging interpretations, and responsibility for the statements made remain with the human researchers. GenAI becomes analytically relevant insofar as its operations alter the way empirical material is integrated into a more comprehensive theoretical account—through condensation, categorization, linking, revision, or integration of the material. This involvement is necessary for the analytical process under investigation but is not sufficient to constitute autonomous theorizing.

Rammert's (2025) concept of gradualized agency within sociotechnical arrangement allows this form of participation to be specified without collapsing technical efficacy into human intentionality. Technical systems can thus be understood as systems that produce effects and respond to changing inputs, their agency arises from the dynamic interplay of causal processes and contingencies. In the present case, the model becomes causally consequential when its outputs redirect the subsequent course of analysis, and contingently responsive when it selects among possible categorizations in relation to interview passages, prompts, previously generated categories, and feedback.

## 2.2 Grounded theory as a contested epistemic practice

GT evolved over time into various strands that emerged, in part, from discussions about how theoretical abstraction should be handled in a way appropriate to the discipline. Even in the original description of the method, Glaser and Strauss had insisted that researchers apply theoretical sensitivity, disciplinary training, professional experience, and existing literature to their data (Glaser and Strauss, 1967). Coding was, contrary to common assumptions, not intended to be a theory-free analysis of raw data.

This insistence has a philosophical background that the later Glaser/Strauss-Corbin dispute tends to obscure. Strauss' own approach to GT draws on what Corbin and Strauss describe as "the tradition of Chicago interactionism and the philosophy of pragmatism inherited largely from John Dewey and George Herbert Mead" (Corbin and Strauss, 2008), thinkers for whom knowledge is not the product of detached observation but is created through (inter)acting. On this view, categories cannot be theory-free in the first place, since the very concepts a researcher brings to the field are themselves products of prior action and interaction, disciplinary and otherwise; the question was never whether analysis is theory-laden, but how that theoretical load is disciplined without foreclosing what the material can still say.

It was this disciplining device, rather than the pragmatist background itself, that became the object of Glaser's critique. What he objected to was something specific: Strauss and Corbin's (1990) coding paradigm, the matrix of causal conditions, context, action/interaction strategies, and consequences used to structure axial coding, which he accused of forcing a predetermined analytic scaffold onto material that should instead be allowed to generate its own categories, a rebuke he titled, tellingly, "Emergence vs. Forcing" (Glaser, 1992).

In Basics of Qualitative Research Strauss and Corbin (1998) explicitly cautioned against applying the coding paradigm rigidly, describing it instead as a guide for provoking thinking about relationships between categories, one that should not prevent researchers from capturing "the dynamic flow of events and the complex nature of relationships" (Strauss and Corbin, 1998).

It was Kathy Charmaz who reframed the dispute rather than resolving it, replacing the discovery-oriented vocabulary both founders had inherited (Charmaz, 2006), a vocabulary literally embedded in the title of their founding text (Glaser and Strauss, 1967), with an explicitly constructivist one. Glaser himself rejected this move, dismissing Charmaz's approach as a departure from GT proper (Glaser, 2007). Mills et al. (2006) describe the resulting methodological landscape less as a resolved dispute than as a continuing "spiral" of positions descending from the same founding text.

Adele Clarke, herself trained by Strauss, pushed the reframing further still, extending rather than abandoning his own social worlds/arenas framework (Clarke, 2005). Her target was not the coding paradigm as such, but GT's residual search for a "pure 'basic social process'" (Clarke, 2005), a core category coherent enough to organize an entire field, which she considered a modernist simplification of what is, in practice, heterogeneous, contradictory, and only partially voiced. Drawing on Foucault, Latour, and Haraway, Clarke (2005) imports into GT a sensitivity to power, materiality, and silence that neither

Glaser's emergence nor Strauss and Corbin's paradigm had been designed to register in this form. This reframing matters well beyond GT's own disciplinary history as it changes what we can (meaningfully) examine when an LLM is instructed to theorize. Against this background, the central question is not whether the model reproduces a theory-free process of discovery, a position that neither Glaser nor Strauss ultimately maintained. Rather, the analytical interest lies in how categories are generated and revised, how contradictory or poorly fitting material is handled, and to what extent the heterogeneity of situations remains visible throughout the process of abstraction (including their material, relational, and often unarticulated dimensions, as emphasized by Clarke (2005)). Otherwise, machine-mediated theorizing might end up reproducing what Clarke (2005) identified as a limitation of conventional GT, which points to the early stabilization of heterogeneous situations into a single, internally consistent representation.

Taken together, Strauss's pragmatist understanding of theory generation, Strauss and Corbin's reformulation of the coding paradigm, Charmaz's (2006) constructivist perspective, and Clarke's (2005) situational approach provide the analytical framework for examining the epistemic work performed within a machine-mediated GT workflow. Rather than determining in advance whether an LLM is capable of theorizing, these perspectives make it possible to reconstruct how categories are generated, revised, and related to one another; whether contradictions and negative cases become analytically consequential; how contextual relations are preserved, transformed, or disappear during abstraction; how categories are eventually stabilized within broader theoretical propositions. They provide the basis for distinguishing between descriptive condensation, categorical organization, and theoretical abstraction without assuming that these operations necessarily coincide.

## 3. Methodology

# 3.1 Research design and analysis

This study examines the epistemic work performed by GenAI when the model is instructed not only to assign codes to empirical passages, but to move through the successive steps of a Grounded Theory workflow by generating codes, categories, relations between them, a core category, and, finally, an integrated theoretical account. The analytical object is therefore neither the interview material as such nor the apparent coherence of the theoretical model produced at the end of this process. What is at stake is the sequence of transformations through which empirical segments are selected, condensed, related, reorganized, and gradually integrated into more abstract categorical structures. Each step has consequences, as it shapes what becomes visible as an analytical relation, what is displaced or lost during abstraction, and also on what grounds the resulting knowledge claims are eventually formulated.

The data corpus consists of analytical artifacts that were produced as outputs during successive phases of the model-mediated research process. These include files containing code generation at the segment level, notes on open coding and a preliminary code book, as well as proposed relationships between codes, axial categories, the selection and justification of a core category, and the theoretical model that

was generated through the model-mediated process. These artifacts neither represent transparent depictions of an internal thought process nor allow for the conclusion that the model, in the methodological sense of GT, independently generated a Grounded Theory. They represent observable indicators which can be used to reconstruct how the model selected and condensed empirical material, established and revised relationships, and gradually stabilized these processes within a broader theoretical framework.

While the model was developed in collaboration with the principal investigator and the project assistant —who has a background in computer science—the analysis was conducted with the entire research team. The team had previously been continuously engaged in ethnographic and analytical work in the field. Their situated knowledge thus shapes the assessment of how certain abstractions should be evaluated in terms of relevant contextual, material, and interactive relationships. The model did not have access to this much broader knowledge of the field, which was based on over 100 interviews and ethnographic fieldwork. Likewise, the existing categories developed by the research team were not used as an evaluation criterion for the machine-generated results. By doing so, the relational interplay between the model's operations and the researchers' situational expertise comes to the fore, highlighting the distributed constitution of analytical judgment within the sociotechnical arrangement.

## 3.2 Corpus and machine-mediated grounded theory workflow

Three interviews (conducted in Turkish) selected from the broader empirical material of the research project were submitted to the model. They represent different organizational positions and forms of knowledge within surgical instrument production, ranging from embodied workshop expertise to engineering, R&D coordination, and firm-level strategic decision-making. This selection introduced analytical variation while keeping the corpus sufficiently delimited to reconstruct in detail how the model moved from coding to category formation and theoretical integration. The corpus was therefore designed for analytical contrast rather than representativeness or theoretical saturation.

The interviews were transcribed (with Transkriptor) and anonymized prior to processing through the API, which divided the 46-, 57-, and 41-page transcripts into 33, 51, and 72 segments. The model was provided with the three interview transcripts (and the methodological instructions embedded in the prompt, but was not supplied with ethnographic fieldnotes, video material, research-team memos, a human codebook, or any categories previously developed within the project. The interviews were conducted in Turkish, whereas the model was instructed to produce its analytical artifacts in English. Translation is therefore not a preparatory step but part of the epistemic work examined here, since terms such as *hissiyat* carry practical connotations that only partially survive the transfer into an English categorical vocabulary.

The workflow was informed by Strauss and Corbin's (1998) distinction between open, axial, and selective coding. During open coding, the model was instructed to analyze the transcripts segment by segment. Then the model was instructed to assign empirically supported codes and to distinguish among

observation, interpretation, theoretical speculation, and methodological uncertainty. It produced coded transcripts, passage-level memos, and a provisional codebook containing both active codes and emergent code candidates.

After open coding, the model had to generate a code-relation network before moving to axial coding. The purpose of this step was to identify recurrent associations and possible clusters across the material without treating co-occurrence as evidence of causality, hierarchy, or theoretical centrality. During axial coding, the open codes were then reorganized in relation to phenomena, conditions, context, action and interaction strategies, and consequences. Selective coding required the model to identify a provisional core category, assess whether it could integrate the axial phenomena, compare it with alternative categories, and develop a theoretical account of the material provided. Contradictory or poorly fitting passages also had to be identified and used to specify the limits of this account. Since the workflow remained restricted to the original three interviews, no theoretical sampling could take place, and the resulting theoretical integration remained limited to this corpus. While open coding was accompanied by passage-level memos, the later stages recorded analytical decisions mainly through category definitions, explanations of relations, and theoretical propositions.

The analysis was performed in July 2026, using OpenAI GPT-5 mini via the OpenAI Responses API (data are not used for model training). Technical settings related to the workflow included a temperature value of 0.0, a maximum output length of 30,000 tokens, no seeding, and a medium reasoning effort setting; responses were requested as JSON Schema structured outputs with strict=False, and API-side response storage was disabled (store=False). Transcripts were segmented according to timestamped speaker returns interviewer and observer returns were retained as local context but excluded from encoding, consecutive returns from the same participant were combined up to a 1,300 character limit and longer returns were split at sentence limits using a minimum target length of 450 characters. These were processed sequentially via interview in API groups containing a maximum of ten analyzable passages or 10,000 characters; this was performed after initializing a temporary codebook from a diversified sample of 36 passages covering the entire dataset. The workflow was executed once. Since a temperature of 0.0 reduces but does not eliminate variation in generated outputs, we treat the analyzed artifacts as one traceable instantiation of the workflow rather than as a reproducible result.

## 3.3 Theory-guided analysis and operationalization

The model-generated artifacts were examined through a deductive, structured qualitative content analysis (Mayring, 2000, Schreier, 2012). Thus, the coding frame was derived from the research question and from the methodological and theoretical discussion developed in Section 2.

Rather than evaluating whether the model correctly applied Grounded Theory, the analysis reconstructed how empirical material was transformed across the successive stages of the model-mediated workflow. Open, axial, and selective coding therefore served to structure the sequence of generated artifacts, but did not constitute the coding framework through which these artifacts were analyzed. Charmaz's (2006) constructivist reformulation and Clarke's (2005) situational critique further sensitized the analysis to the

situated character of categories and to the treatment of heterogeneity, contradiction, materiality, and what remains unarticulated.

The coding unit was a discrete analytical operation performed in the generated outputs. This could consist of transforming an interview passage into code, integrating several codes into an axial category, constructing a relation between categories, selecting a core category, or formulating a theoretical proposition. The context unit comprised the original interview passage and its immediate textual context, together with the associated code, memo, category definition, relational justification, or theoretical statement. The unit of analysis was the complete traceable sequence through which a particular empirical element was carried from the interview material into a more abstract analytical form. It was possible to assess not only the plausibility of the final output, but also the transformations through which it was produced by distinguishing between these units (Mayring, 2000, Schreier, 2012).

The analysis was conducted along six interrelated criteria, which made it possible to reconstruct the transition from the empirical material to theoretical integration. The first was empirical grounding, i.e., whether a code, a category, or a theoretical statement could be traced back to the passages used to support it, or whether the analytical claim went beyond what these passages could substantiate. The second concerned conceptual transformation. This referred to the changes that occurred as the model shifted from empirical description to analytical abstraction: what it selected and condensed, as well as what it generalized, added, reformulated, or shifted during this process. The third aspect was relational adequacy. Here, we looked both at whether codes were linked to one another and whether the nature and direction of the proposed relationship (sequential, causal, conditional, or reciprocal) could be supported by the data. Fourth, we evaluated the response to contradictions. This involved the question of what happens when text passages cannot be fitted into an emerging category. Did this lead to their revision, differentiation, or restriction, or did the system integrate them in such a way that the category itself remained unchanged? The fifth dimension concerned contextual sensitivity. Here, we examined the extent to which the positions of the actors, temporal sequences, organizational frameworks, material relationships, and conditions of interaction remained analytically relevant rather than disappearing in the process of abstraction. The final dimension concerned theoretical integration—that is, whether the relationships established between the categories made empirical variations understandable, or whether recurring themes were primarily condensed into a coherent but largely descriptive account.

These dimensions were operationalized through definitions and decision rules, thereby making the basis for evaluation explicit without reducing the analysis to a formal evaluation procedure. An empirical argument was considered well-supported if the cited passage directly confirmed the analytical argument and partially supported if the argument depended on additional interpretive steps or a combination of material from different contexts. Similarly, a proposed relationship was considered sufficiently substantiated only if the material confirmed both the connection and its direction, since the mere co-occurrence of codes is not sufficient to determine whether one phenomenon precedes, conditions, produces, or modifies another. Addressing contradictions required more than merely identifying a negative or non-conforming case. It became analytically relevant only when contradictory material led to

a verifiable revision, qualification, differentiation, or rejection of the emerging category. When examining whether the resulting account specified mechanisms, conditions, variations, and consequences—rather than simply summarizing recurring themes under a common conceptual heading—a distinction was made between theoretical integration and thematic abstraction.

From the complete set of generated artifacts, five diagnostic cases were selected. This selection followed the analytical sequence of the workflow and was not based on the frequency of particular codes or on the apparent persuasiveness of the outputs alone. Rather, the cases were chosen because they made different forms of epistemic operation observable across the successive stages of the process, including code generation, category formation, the construction of relations, selective integration, and the treatment of contradiction. Their complete analytical trajectories could be reconstructed across the generated files, while the differences between them in analytical scale and empirical support made it possible to examine how the model's operations changed as the workflow moved towards increasing levels of abstraction.

For each case analysis, the relevant (coded) interview segment, the corresponding model-generated artifacts, and their successive development throughout the entire workflow were tracked and examined with reference to the six analytical dimensions outlined above. The following analysis therefore highlights those dimensions that were of particular significance. The assessments addressed the extent to which the generated artifacts were considered empirically supported, distinguishing between well-founded and partially well-founded interpretations. These classifications served as analytical assessments and not as objective quality ratings. The assessments were made by the first author and discussed within the research team; contested cases were resolved through consensual coding rather than through measures of intercoder agreement. The aim was not reliability in a metric sense but intersubjective traceability, which is why decision rules and artifact references are documented throughout.

## 4. Results

# 4.1 From empirical passage to analytical code

The first case discusses an interview passage in which an experienced craftsman describes his transition from instrument repair to production. The interview partner explains that he would have preferred to learn alongside an experienced master. However, when he joined the company, the craftsmen who had previously carried out this work had already moved into subcontracting, leaving an internal gap in skilled production knowledge. After nine months in instrument repair, he was transferred to production and asked to manufacture an instrument he had never seen or made before. He recalls learning through repeated attempts and failures:

"We made one and ruined three. By continuing like this, I eventually reached the point where there is almost no type of instrument I cannot make." (TR03MasF1, orig. *Bir tane yaptık, üç tane bozduk. Öyle diye*

*diye. Şu anda yapmadığım, çeşit olarak, alet hemen hemen yoktur yani.*)

For this claim, the model proposed *Learning-by-trial-and-error / error-prone practice* as an emergent code (Axial Coding Phase.docx). The corresponding memo related the participant's repair experience to repeated production failures and the gradual development of practical competence. It also suggested that repair and learning through error may have compensated for the absence of formal training (TR03MasF1_open_coding_memos.docx).

The proposed code is closely connected to the passage, since the participant himself relates repeated production failures to the gradual acquisition of competence (TR03MasF1_open_coded.docx, TR03MasF1_phase1_open_coding.xlsx, EC10[1]). At the same time, the passage contains further relations that are not captured by the code itself. It refers to the departure of experienced masters, damaged instruments, rejection by quality control, manual aptitude, and the emotional strain of producing an instrument without prior demonstration (TR03MasF1_open_coded.docx, TR03MasF1_p0014–p0016). The code foregrounds one relation within this account, namely the movement from repeated failure to competence, while other organizational, material, and affective dimensions are either registered in the corresponding and adjacent memos or remain outside the categorical formulation (TR03MasF1_open_coding_memos.docx, TR03MasF1_p0014 and TR03MasF1_p0016; TR03MasF1_phase1_open_coding.xlsx, EC10).

We therefore assessed the proposed code as substantiated, whereas the interpretation that trial and error substituted for apprenticeship-based guidance was only partially supported. The passage establishes a relation between error, repetition, and acquired competence, but does not by itself constitute practical learning and formal training as general alternatives (TR03MasF1_open_coded.docx, TR03MasF1_p0016; TR03MasF1_open_coding_memos.docx, TR03MasF1_p0016). The status of the code within the workflow is also relevant. EC10 remained an emergent candidate and not an active code in the finalized codebook (TR03MasF1_phase1_open_coding1.xlsx, EC10; Phase 1 Summary.docx, "Unresolved Codebook Issues" and "Warnings"). But code EC10 was subsequently incorporated into the axial category *Tacit craft knowledge and skill formation* (Axial Coding Phase.docx, PH3, supporting code EC10; TR03MasF1_p0016). Its provisional status did not disappear formally. However, it became less consequential as the analysis proceeded. This is where the first movement of stabilization becomes visible.

# 4.2 From codes to axial coding and building an axial category

The axial category *Tacit craft knowledge and skill formation* includes the absence of an experienced master, learning by trial-and-error, prior knowledge of instruments through repair, embodied *hissiyat (engl. feeling)*, and the time required to acquire professional competence. In the axial coding phase, the model arranged these elements as a process in which mentor absence created conditions for informal learning, repeated failure generated tacit competence, and professional mastery developed over several

years (Axial Coding Phase.docx, PH3, “Paradigmatic Model,” “Causal Conditions,” “Context,” and “Action and Interaction Strategies”).

Each element can be traced to explicit passages of the interview with the craftsman, who describes how the departure of the masters left a gap in the workshop, recalls that “we made one and ruined three.” TR03MasF1 then explains that his previous work repairing and sharpening instruments provided a practical basis for production, followed by the completion of an instrument as a matter of what he describes as *hissiyat* that cannot simply be seen and estimates that independent production requires at least three years and mastery approximately five (TR03MasF1_open_coded.docx, TR03MasF1_p0014, p0016, p0018, p0081, and p0083; TR03MasF1_open_coding_memos.docx). The category *Tacit craft knowledge and skill formation* therefore connects an organizational condition to a mode of learning, an embodied form of judgment, and a temporal account of skill formation.

The elements are described as interconnected parts of the participant’s professional career. However, this does not represent a continuous causal chain in which the absence of a mentor inevitably leads to experimentation, and experimentation in turn produces *hissiyat*. The axial analysis thus interprets episodes of trial and error as the cause of the acquisition of tacit skills and the absence of a mentor as creating the conditions for on-the-job learning. Furthermore, the category is linked to the assignment of professional roles, the preference for hiring younger workers, and the longevity of products and craft practices (Axial Coding Phase.docx, PH3, “Causal Conditions,” “Action and Interaction Strategies,” and “Consequences”; TR03MasF1_p0034, pp. 0063 and 0187).

For this reason, the category itself can be interpreted as empirically substantiated, even if its internal causal sequence is incomplete. It is based primarily on a single interview. The more comprehensive corpus contains further references to practical craftsmanship, knowledge of materials, gaps in mentorship, and staff training, but these do not reflect the same sequence in comparable depth (Axial Coding Phase.docx, PH3, “Intervening Conditions”; Phase 1 Summary.docx, “Strongest Recurring Concepts” and “Staff Training, Lack of Mentors, and Hiring Procedures”).

Upon closer examination, the category also shifts the level at which competence development is explained. *Hissiyat* appears primarily as a competence that the individual worker has acquired over time. Tools, materials, repair practices, quality control measures, workshop hierarchies, and the relationships through which workers observe, correct, and evaluate one another are still present in the empirical passages, but they recede into the background within the axial category. The model thus provides a process-oriented account of competence development, with the focus on the individual artisan rather than on the broader socio-material framework through which practical judgment emerges.

## 4.3 Constructing relations across the material

The third case involves the construction of a process connecting clinician or customer demands, R&D, technical drawing, production, quality testing, and possible series production. The model brought together passages in which feedback from surgeons was transmitted to and examined by R&D,

customers initiated development projects, workshop personnel collaborated with R&D on technical drawings, and parts were subsequently produced through manual or CNC-supported processes. Other passages affected functional testing, revision after clinical assessment, and the possibility that an accepted prototype could enter series production (TR03MasF1_open_coded.docx, TR03MasF1_p0043, p0113, p0123, p0125, and p0127; TR12ExpF1_open_coded.docx, TR12ExpF1_p0060). The axial output integrated these elements into the following sequence:

“clinician or customer demand → R&D evaluation and prototyping → technical drawing or sample → manual or CNC-supported production → quality testing → series production, special production, revision, or rejection.” (Axial Coding Phase.docx, PH1, “Paradigmatic Model,” “Relations to Other Main Phenomena,” and “Preliminary Analytical Memo”).

The individual components of this process are substantially supported. The interviews show that user feedback can reach R&D, that customers may bring projects to the company for development, and that workshop and engineering personnel cooperate in translating a product idea into drawings and production steps. TR03MasF1 and TR12ExpF1 also document CNC processing, internal functional testing, iterative revision, and the possibility that approval of the surgeon may enable further production (TR03MasF1_open_coded.docx, TR03MasF1_p0039, p0113, p0123, p0125, and p0127; TR12ExpF1_open_coded.docx, TR12ExpF1_p0060). The system had already identified a “strong processual chain” in Phase 1 and connected user feedback, R&D adaptation, technical drawings, production, quality testing, and catalog or series outcomes (Phase 1 Summary.docx, “Initial Analytical Observations”). The analysis in the axial coding further positioned restricted access to operating rooms, prototype status, regulatory documentation, and technical feasibility as conditions that could enable, modify, or interrupt this process (Axial Coding Phase.docx, PH1, “Context,” “Intervening Conditions,” and “Contrasting and Negative Cases”; TR12ExpF1_open_coded.docx, TR12ExpF1_p0008, p0089, and p0098; TR03MasF1_open_coded.docx, TR03MasF1_p0093).

The model linked general descriptions of quality control procedures to specific development phases, while technical documentation, clinical access, and marketing channels became integral parts of the same process. From this, it can be concluded that the model linked parts of the different interviews, organizational positions, surgical products, and production phases (which were not presented as a single, coherent process) with one another. In doing so, it created continuity between empirically interconnected practices that had not necessarily been organized into a single, consistently institutionalized chain.

We then compared the axial analysis with the previous network analysis output. In Phase 1.5, 71 detected relationships were still classified as uncertain and undirected co-occurrences. The output explicitly cautioned against interpreting the results as causal or process-related without further synthesis and confirmation by the researchers (Phase 1.5 Network Analysis. docx, “Quality Control,” “Warnings,” and “Next Required Input”). During axial coding, several of these associations were transformed into causal conditions, strategies, intervening conditions, and consequences (Axial Coding

Phase.docx, PH1). Some of these conversions are supported by directional cues in the interviews, e.g., when user feedback triggers research and development activities or when the assessment by healthcare professionals leads to a revision. Others are based on the integration of text passages into the model whose context is plausible but which are not themselves described as directional.

The output's audit trail also contained a specific attribution error. The axial analysis assigned the account in which a clinician generally approved an instrument but returned certain parts for revision to TR03MasF1_p0133 (Axial Coding Phase.*docx*, PH1, "Consequences"). In the open-coding output, this event is documented under TR03MasF1_p0127, whereas p0133 concerns the clinician's subsequent visit and the joint examination of the instrument (TR03MasF1_open_coding_memos.docx, TR03MasF1_p0127 and p0133). The empirical event is therefore available, but the reference through which the axial relation is made traceable is inaccurate.

The individual relationships can be considered well-supported, but the integrated sequence is only partially supported. During the axial coding phase, the model reveals a distributed organizational process by linking clinical requirements to R&D, production, testing, and potential product outcomes. At the same time, it translates varying and sometimes discontinuous practices into a comparatively linear chain. This becomes visible at the point where prototype approval, catalog inclusion, pricing, and mass production are linked together, even though the findings of the model state that the institutional criteria governing these transitions remain insufficiently documented (Axial Coding Phase.docx, PH1, "Open Analytical Questions," "Completeness of the Paradigmatic Model," and "Methodological Uncertainty").

## 4.4 Selecting and stabilizing a core category

During selective coding, the model generated the core category *"Clinician-mediated productification"*, which is described as a process in which requirements are incorporated into R&D, translated into prototypes or technical drawings, and subsequently transferred to manual or CNC-assisted manufacturing. Following testing and revisions, a surgical instrument may be added to the catalog or enter series or custom manufacturing. Regulatory requirements, delivery capacities, quality control, and staff qualifications were identified as conditions that shape this process (Phase 2 Selective Coding and Rejected Categories.docx, "Provisional Core Category" and "Central Process").

The term *"clinician-mediated"* was discussed within our team, because the relevant interviewees primarily refer to surgeons who formulate requirements, evaluate instruments, and influence decisions regarding revisions or further production. The term *"clinician"* broadens this role to include a wider group of healthcare professionals who are not represented in the data in this specific way. The choice of this term was not re-examined by the model in subsequent phases. As a result, this category was adopted as is, and subsequent relationships were organized around it (*Phase 2 Selective Coding and Rejected Categories.docx*, "Provisional Core Category"; *Theory Construction.docx*, "Core Category" and "Category Relationships"). Through repeated analytical use, a provisional label thus became part of the theoretical structure.

The model compared this category with three alternatives: *Shop-floor production chain and quality governance*, *Tacit craft knowledge and skill formation*, and *Structural market and regulatory constraints on production*. Shop-floor production was documented, tacit knowledge was well-supported in relation to learning and quality judgment, and structural constraints appeared across the interviews. They were nevertheless subordinated because, in the model's assessment, they did not connect an initiating actor, an organizational process, and a product outcome with the same reach (*Phase 2 Selective Coding and Rejected Categories.docx*, "Selection Justification" and "Rejected Core Category Candidates"; *final.docx*, sections 5–6).

The model provides a clear insight into the selection criteria for the core category by showing that alternatives were compared and the remaining categories were reorganized around the selected core. However, when considering all three interviews, the central role of the core category is not convincingly evident. The relationship between surgeons, R&D, and production is most clearly documented by TR03MasF1 and TR12ExpF1. TR33ExpF6 mainly contains historical accounts of patents, international competition, export orientation, and the withdrawal from the manufacturing of surgical instruments. These materials cover the same industrial sector but only support the core category after they are reinterpreted as conditions, variations, or consequences of productification (*Phase 2 Selective Coding and Rejected Categories.docx*, "Variation Across Dataset," "Negative Cases and Scope Limits," and "Methodological Uncertainty").

We therefore regarded the selection procedure as substantiated, but the centrality attributed to the category as only partially supported. The choice privileges a process with a recognizable starting point and a product-related outcome. Embodied knowledge, shop-floor governance, and market relations remain part of the analysis, but no longer appear as possible centres of explanation. Their relevance is determined by the position assigned to them within productification.

The subsequent theory construction reinforced this hierarchy. The model positioned shop-floor production and quality control as the subsequent stage in the clinician-mediated productification process. This relation was supported primarily by passages describing the transition from collaborative R&D drawing to CNC-based production (TR03MasF1_p0113; TR03MasF1_p0123), rather than by the broader set of passages documenting shop-floor organization itself. Tacit knowledge was linked reciprocally to production and R&D (Theory Construction.docx, "Category Relationships" and "Analytic Model"). The evidence for this last relation is weaker. The cited passages document *hissiyat*, trial-and-error learning, and the time required to become a skilled practitioner. They do not show directly that R&D formalizes tacit knowledge or that craft expertise reorganizes R&D decisions. The references make the categories traceable, but do not fully substantiate the direction assigned to their relation.

## 4.5 Integrating contradiction into a theoretical account

The final case concerns two findings that sit uneasily with the selected core category. TR12ExpF1 explains that invitations to operating rooms had become less frequent because infection risks and patient-safety concerns made access more restrictive, thereby reducing opportunities for direct

observation of surgical practice (TR12ExpF1_open_coded.docx, TR12ExpF1_p0089–p0090; TR12ExpF1_open_coding_memos.docx, TR12ExpF1_p0089–p0090). TR33ExpF6 explained that the firm had ended surgical hand-instrument production around fourteen years earlier. It had not left the medical sector, but shifted towards four medical-device groups: surgical motor systems, sterilization equipment, medical lighting systems, and disposable catheters (TR33ExpF6_open_coded.docx, TR33ExpF6_p0002–p0003 and p0009–p0010).

The model retained both findings. Restricted operating-room access was classified as a contradiction during axial coding and retained under “Negative Cases and Scope Limits” in the subsequent selective-coding output (Axial Coding Phase.docx, “Contrasting and Negative Cases”; Phase 2 Selective Coding and Rejected Categories.docx, “Negative Cases and Scope Limits”). The supporting passage, TR12ExpF1_p0089, describes how infection-control concerns had reduced invitations to observe in operating rooms. In both outputs, the model interpreted this restriction as weakening the clinician-observation channel and limiting the clinician-mediated process (Axial Coding Phase.docx, “Contrasting and Negative Cases”; Phase 2 Selective Coding and Rejected Categories.docx, “Negative Cases and Scope Limits”).

In the subsequent theory construction, the same passage, TR12ExpF1_p0089, was incorporated into the paradigmatic model under “Intervening Conditions” as operating-room observation access and infection-related restrictions (Theory Construction.docx, “Intervening Conditions”). At the same time, it remained listed under “Negative Cases and Tensions,” where restricted observation access was connected to reduced field-based feedback for R&D (Theory Construction.docx, “Negative Cases and Tensions”).

The withdrawal from surgical hand-instrument production was initially identified as an alternative process. The relevant passage, TR33ExpF6_p0002, states that the firm had ended its involvement in surgical-instrument production approximately fourteen years earlier (Phase 2 Selective Coding and Rejected Categories.docx, “Negative Cases and Scope Limits”). In the theory-construction output, this strategic exit was subsequently positioned among the principal consequences of regulatory, market, and competitive conditions and was also retained as a source of heterogeneity among firms (Theory Construction.docx, “Principal Consequences” and “Negative Cases and Tensions”).

The text provides an example of restricted access to operating rooms, from which the model inferred that opportunities for observation and clinical exchange had diminished (Phase 2 Selective Coding and Rejected Categories.docx, “Negative Cases and Scope Limits,” TR12ExpF1_p0089). However, the consequences of this restriction are not demonstrated through a specific product-development case. Nevertheless, the model incorporated the same restriction into the paradigmatic model as an intervening condition that reduced feedback from practice to R&D (Theory Construction.docx, “Intervening Conditions” and “Negative Cases and Tensions,” TR12ExpF1_p0089).

The company’s shift in focus from the manufacturing of hand instruments to other medical technologies was attributed to limited export prospects, its competitive position, and a strategic decision (TR33ExpF6_open_coded.docx, TR33ExpF6_p0002–p0003; Theory Construction.docx, “Conditions

Shaping Variation” and “Principal Consequences”). It was not narrated as the failed outcome of a specific surgeon-initiated project. The firm’s later export success also concerned the transformed product portfolio rather than a continuation of its hand-instrument production (TR33ExpF6_open_coded.docx, TR33ExpF6_p0187).

This distinction matters because the presence of a passage reference makes an interpretation traceable without necessarily substantiating the mechanism attributed to it. In the cases examined here, the model connected empirically documented processes through broader theoretical relations that were not themselves described by the participants. Restricted operating-room access was directly documented, but its incorporation as a condition reducing practice-based feedback to R&D represented a further interpretive step (Phase 2 Selective Coding and Rejected Categories.docx, “Negative Cases and Scope Limits,” TR12ExpF1_p0089; Theory Construction.docx, “Intervening Conditions” and “Negative Cases and Tensions”). Similarly, the company’s withdrawal from hand-instrument production was empirically documented, whereas its positioning as a consequence of regulatory, market, and competitive conditions was produced through the subsequent theoretical integration (TR33ExpF6_open_coded.docx, TR33ExpF6_p0002–p0003 and p0187; Theory Construction.docx, “Conditions Shaping Variation” and “Principal Consequences”).

Neither case substantially altered the core category. Instead, restricted access was incorporated simultaneously as an intervening condition and as a continuing tension, while strategic withdrawal was positioned as an alternative process, a source of variation, and a possible consequence (Phase 2 Selective Coding and Rejected Categories.docx, “Negative Cases and Scope Limits”; Theory Construction.docx, “Intervening Conditions,” “Principal Consequences,” and “Negative Cases and Tensions”). We use categorical absorption to describe this operation. Material that could unsettle the organizing category remains visible, but is translated into analytical positions that allow the central process to persist.

Across the five cases, the model moved from passage-related condensation towards progressively broader integration. Emergent code candidates were incorporated into axial categories, relations initially recorded without a specified direction were subsequently rendered as sequential or conditional, alternative categories were assigned subordinate functions, and contradictory material was retained within the theoretical model (Preliminary Codebook.docx, “Emergent Code Candidates”; Code Relation Network.docx; Axial Coding Phase.docx; Phase 2 Selective Coding and Rejected Categories.docx; Theory Construction.docx). The resulting account is more than a thematic summary. Yet its theoretical productivity rests on a successive stabilization through which provisional relations are increasingly treated as components of an integrated account. This produces coherence, but also makes it more difficult to distinguish empirically established relations from those generated through the model’s own theoretical integration, thereby reducing the visibility of other ways in which the material might have been organized.

# 5. Discussion

As mentioned in the introduction, AI-supported qualitative research primarily focuses on coding accuracy or compares between humans and machines. This study distances it from this dichotomous perspective with its focus on the epistemic work performed within the framework of an LLM-supported GT workflow. We have demonstrated how theoretical relationships emerged, consolidated, and gained analytical authority through successive interactions between machine-generated results and researchers' verification. Machine-assisted theory-building is best understood as a sociotechnical arrangement rather than as the product of either autonomous computation or human interpretation alone.

In Section 4, we showed that the model performed operations that went beyond summarizing text passages. It generated concept candidates, organized categories, specified relationships, and proposed an integrated theoretical explanation. These operations are methodologically interesting because they structured what was considered the central analytical process. At the same time, they remain embedded in a workflow shaped by the selection of interviews, the design of prompts, grounded theory procedures, the computational infrastructure, and the researchers' interpretive decisions. In line with the relational understandings of agency (Orlikowski 2007; Suchman 2007, 2011; Rammert 2025), the model played a defining role in theory-building, without, however, independently determining the research question or the epistemic standards by which its findings were evaluated.

The workflow reproduces some of the conditions associated with "forcing" in Glaser's (1992) critique. The axial coding prompt instructed the model to organize the material through phenomena, conditions, strategies, and consequences, thereby providing a relational structure within which passages could be integrated. The transformation of initially undirected relations into causal or conditional ones can therefore not be separated from the methodological architecture imposed by the workflow.

The findings show that the problem does not end with the direct application of a coding paradigm. Most individual transitions remained plausible when considered on their own. The more consequential movement occurred across phases, as provisional codes, relations, and category labels became the basis for subsequent analytical operations. We describe this cumulative process as *recursive stabilization*. Forcing and recursive stabilization are therefore not competing explanations. The former concerns the structuring effects of the methodological scheme, while the latter captures how early analytical decisions acquire increasing authority as they are inherited by later phases.

The term "*clinician*," e.g., extends the empirically specific reference to surgeons to a broader group of healthcare professionals. Once introduced, the term was incorporated into theory-building without its empirical scope being reexamined. What thus became stabilized was not only a series of relationships but also a category that was broader than the underlying material justified. These transitions are not socially neutral. The forms of knowledge in our corpus did not enter the abstraction on equal terms: accounts formulated in organizational and product-oriented vocabulary were readily converted into conditions, strategies, and consequences, whereas embodied workshop knowledge, such as the *hissiyat* acquired through years of repetition, entered the theoretical account mainly as a precondition of

production rather than as expertise in its own right. This asymmetry also has a linguistic dimension. In the early coding phases, the model occasionally retained Turkish terms as code labels, whereas later phases operated almost exclusively in an English categorical vocabulary. What survived this shift were concepts that could be rendered in the organizational idiom of the workflow, while terms whose practical connotations resist translation—*hissiyat* being the clearest case—were carried forward only in attenuated form.

At the level of a single coding decision, imprecise labels such as 'clinician' may seem minor. However, they gain significance when subsequent relationships are organized around the adopted category. Recursive stabilization thus operates in two interrelated ways: provisional relationships become increasingly difficult to distinguish from well-founded ones, while early conceptual generalizations gain theoretical weight through continued use. Conversely, this means that the recursive stabilization phase requires a human actor to reexamine established categories.

The difference between traceability and substantiation also becomes clear in this context. Passage identifiers, citations, and audit trails make machine-generated analytical decisions traceable. However, they do not prove that the material supports the theoretical relationship later attributed to it. Although limited access to operating rooms and the strategic withdrawal from the production of surgical instruments were empirically documented, they were later integrated into broader processual relations that the participants themselves had not described. While traceability facilitates methodological verification, the question of whether the relationship remains sufficiently grounded in the material must still be assessed by a human.

These findings have far-reaching implications for AI-supported qualitative research. The involvement of humans should not be understood primarily as a final verification step that corrects machine-generated errors. Instead, it involves critically examining the analytical transitions through which preliminary categorizations acquire theoretical status. This includes the question of when emerging categories become established, when co-occurrences are transformed into processual relationships, and when contradictory material is integrated without altering the emerging theoretical core. In practice, researcher review should therefore be organized around these analytical transitions.

Because the analysis concerns one GT workflow and a small corpus from surgical instrument manufacturing, further research is needed to examine whether recursive stabilization also occurs in other qualitative methods, model architectures, languages, and empirical domains, and whether it varies across repeated runs of the same workflow.

The implications of GenAI for qualitative social research emerge from the interplay of methodological conventions, sociotechnical infrastructures, and distributed epistemic work. GT offers one productive perspective, as it reveals how categories are successively generated, revised, stabilized, and sometimes protected from revision. The challenge for future AI-supported qualitative research therefore lies in understanding how its theoretical claims are stabilized, evaluated, and, where necessary, disrupted within research practice.

# Declarations

Author contributions

M.Ş. conceptualized the article, developed the research question and analytical framework, and led the methodological design of the study. M.Ş. and G.B. conducted the interviews that provided the empirical material for the analysis. A.B.U. contributed to the selection of cases and the preparation of the interviews.

M.Ş. and U.K. jointly developed the prompt and computational workflow, with M.Ş. determining the principal analytical and methodological specifications. U.K. implemented the API-based software, executed the computational workflow, and tested, debugged, and modified the program where technical adjustments were required.

M.Ş. reviewed the generated outputs, conducted the detailed qualitative analysis, interpreted the findings, and developed the article's central theoretical and methodological arguments. G.B. and A.B.U. contributed to the discussion and interpretation of selected findings and to the conceptual refinement of the operationalization of epistemic work.

M.Ş. wrote the original draft of the manuscript and integrated the subsequent revisions. All authors critically reviewed and discussed the final manuscript, approved the submitted version, and accept responsibility for their respective contributions.

Ethical approval

The MediCluster research project was reviewed and approved by the Ethics Committee of the Max Weber Stiftung (Max Weber Foundation) on 21 July 2025 (Approval Reference: 2025-03 MediCluster Sahinol OII). The committee confirmed that the ethical issues associated with research involving human-participant data had been identified and appropriately addressed. The research was conducted in accordance with the ethical requirements specified in the approval and with the relevant institutional guidelines and regulations.

Informed consent

Informed consent was obtained from all participants prior to their participation in the study. Participants were informed about the purpose of the research, the voluntary nature of participation, the handling and protection of their data, and their right to withdraw in accordance with the conditions communicated during the consent process.

Competing interests

The authors declare no competing interests.

Data availability

The qualitative data generated and analysed during this study, including interview transcripts and associated field materials, are not publicly available because they contain sensitive, context-rich information and because public disclosure could compromise the confidentiality of participants and organizations, even after de-identification.

Access to limited and appropriately de-identified research data may be considered by the corresponding author upon reasonable request. Any access is subject to the scope of the participants' informed consent, applicable data-protection requirements, approval by the responsible institutional and ethical bodies, and, where appropriate, the conclusion of a data-use agreement. Full interview transcripts and materials that could enable the direct or indirect identification of participants, organizations, or research sites will not be shared.

Availability of materials and code

The system prompt and the methodological specifications used for the computational analysis are provided in the Supplementary Information. The custom code used to implement and execute the API-based analytical workflow is available from the corresponding author upon reasonable request. Access may be subject to technical, licensing, security, or data-protection restrictions.

Machine-generated outputs containing quotations or other information derived from the interview material are subject to the same ethical and confidentiality restrictions as the underlying qualitative data.

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project number 551920228.

## References

1. BIANCO A, BROCCA N, GARASSINO D (2026) Assessing the Potential of LLM-assisted Annotation for Corpus Pragmatics: The Case of Humor. Corpus Pragmat, 10
2. BREAZU P, SCHIRMER M, HU, S., KATSOS N (2024) Large Language Models and the challenge of analyzing discriminatory discourse: human-AI synergy in researching hate speech on social media. J Multicultural Discourses 19:157–175
3. CARMONA-DíAZ G, JIMéNEZ-LEAL W, SRIPADA GRISALESMA, AMAYA C, INZLICHT S, M., BERMúDEZ JP (2026) An AI-powered research assistant in the lab: A practical guide for text analysis through iterative collaboration with LLMs. Behav Res Methods, 58
4. CHARMAZ K (2006) Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis. Sage, Thousand Oaks, CA
5. CLARKE AE (2005) Situational Analysis: Grounded Theory After the Postmodern Turn. Sage, Thousand Oaks, CA
6. COOK SD, BROWN JS (1999) Bridging epistemologies: The generative dance between organizational knowledge and organizational knowing. Organ Sci 10:381–400

7. CORBIN J, STRAUSS A (2008) Basics of qualitative research: Techniques and procedures for developing grounded theory, 3rd edn. Sage, Thousand Oaks, CA
8. GLASER BG (1992) Basics of Grounded Theory Analysis: Emergence vs. Forcing. Sociology, Mill Valley, CA
9. GLASER BG (2007) Constructivist grounded theory? Hist Social Research/ Historische Sozialforschung 19:93–105
10. GLASER BG, STRAUSS AL (1967) *The Discovery of Grounded Theory*: *Strategies for Qualitative Research*, Chicago, Aldine
11. HAN Z, TAVASI A, LEE J, OPPENHEIM LUZURIAGAJSURESHK, BATTAGLIA M, F., TERLECKY SR (2025) Can large language models be used to code text for thematic analysis? An explorative study. Discover Artif Intell, 5
12. HODGES A, PONCE T, SEAWRIGHT L (2025) AI-Assisted Qualitative Coding: Comparing Human to Machine Outputs. IEEE International Professional Communication Conference, Institute of Electrical and Electronics Engineers Inc., 230–233
13. JAYAWARDANA BC, WITHANAARACHCHI AS, JAYALAL S, ABEYSEKARA R (2026) Evaluating NVivo, R, and Large Language Models for Thematic Analysis: Comparative Computational Insights for Qualitative Researchers. 2026 6th International Conference on Advanced Research in Computing: Responsible AGI: Balancing Intelligence, Responsibility and Sustainability, ICARC 2026 - Conference Proceedings, Institute of Electrical and Electronics Engineers Inc
14. KHALID MT, WITMER AP (2025) Prompt Engineering for Large Language Model-Assisted Inductive Thematic Analysis. Social Sci Comput Rev
15. KNORR-CETINA KD (1999) Epistemic Cultures: How the Sciences Make Knowledge. Harvard University Press, Cambridge, MA
16. MAYRING P (2000) Qualitative Inhaltsanalyse. Forum Qualitative Sozialforschung/Forum: Qualitative Social Research. Online J, 1
17. MILLS J, BONNER A, FRANCIS K (2006) The development of constructivist grounded theory. Int J qualitative methods 5:25–35
18. ORLIKOWSKI WJ (2007) Sociomaterial practices: Exploring technology at work. Organ Stud 28:1435–1448
19. PARK JK, ELLEZHUTHIL RD, WISNIEWSKI P, SINGH V (2025) Collaborative human-AI risk annotation: co-annotating online incivility with CHAIRA. Inform Res 30:992–1008
20. RAMANATHAN S, LIM LA MOTTAGHI, N. R., SHUM, S. B. & ACM 2025. When the Prompt becomes the Codebook: Grounded Prompt Engineering (GROPROE) and its application to Belonging Analytics. *FIFTEENTH INTERNATIONAL CONFERENCE ON LEARNING ANALYTICS & KNOWLEDGE, LAK 2025.*
21. RAMMERT W (2025) Technology and the Theory of Distributed Agency. Distributed Agency & Digital Technology: A Social Pragmatist View on Human-Technology Interaction. Springer Fachmedien Wiesbaden, Wiesbaden

22. SCHREIER M (2012) Qualitative Content Analysis in Practice. First Edition ed. 55 City Road. SAGE Publications Ltd.
23. SHARMA A, & WALLACEJR (2025) ACM DeTAILS: Deep Thematic Analysis with Iterative LLM Support. *PROCEEDINGS OF THE 2025 CONFERENCE ON CONVERSATIONAL USER INTERFACES, CUI 2025*.
24. STRAUSS A (1985) Work and the division of labor. Sociol Q 26:1–19
25. STRAUSS A (1988) The articulation of project work: An organizational process. Sociol Q 29:163–178
26. STRAUSS A, CORBIN J (1990) *Basics of qualitative research*, sage Newbury Park, CA
27. STRAUSS A, CORBIN J (1998) Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Sage, Thousand Oaks, CA
28. STRAUSS AL (1993) *Continual permutations of action*, AldineTransaction
29. SUCHMAN L (2007) Human–Machine Reconfigurations: Plans and Situated Actions. Cambridge University Press, Cambridge
30. SUCHMAN L (2011) Anthropological relocations and the limits of design. Annu Rev Anthropol 40:1–18
31. THOMINET L, ACOSTA AMORIMJ, K., SOHAN VK (2024) Role Play: Conversational Roles as a Framework for Reflexive Practice in AI-Assisted Qualitative Research. J Tech Writ Communication 54:396–418
32. XIAO Z, YUAN XD, LIAO QV ABDELGHANI, R., OUDEYER, P. Y. & ACM 2023. Supporting Qualitative Analysis with Large Language Models: Combining Codebook with GPT-3 for Deductive Coding. *COMPANION PROCEEDINGS OF 2023 28TH ANNUAL CONFERENCE ON INTELLIGENT USER INTERFACES, IUI 2023 COMPANION*.

# Footnotes

1. The prefix “EC” was used by the computational workflow as a technical identifier for an emergent code or code candidate generated during open coding. It does not denote a recognized GT coding category, and its precise expansion was not specified in the system prompt.