=Draft chapter from Baker, Ryan S.; Barany, Amanda (forthcoming) Artificial Intelligence in Qualitative Research: New Possibilities. Philadelphia, PA: Vassant Press.

Find the latest chapters online at
https://learninganalytics.upenn.edu/book/ai-qualitative-research.html

Cite this chapter as Baker, Ryan S.; Barany, Amanda (forthcoming) Artificial Intelligence in Qualitative Research: New Possibilities. Ch. 5: A Workflow for the Qualitative Research Process. Philadelphia, PA: Vassant Press.

Chapter 5. A Workflow for the Qualitative Research Process
Version 1.02, 21 July 2026

Since the beginning of formal qualitative research, scholars have formed communities who choose and negotiate the goals, contexts of inquiry, and priorities that guide their work. Understanding and aligning our work with these ways of thinking about and doing research (sometimes called research paradigms) can help us make thoughtful planning decisions for a research study, including:

(1) What kinds of **questions** are worth asking?

(2) What **information** do I need to answer these questions?

(3) What **processes** will help me turn that information into answers?

The design of a research study is shaped by a series of decisions about framing, data, and interpretation that change and accumulate over time (Creswell & Poth, 2018; Ravitch & Carl, 2021). Research methods literature often introduces us to the specific traditions in these schools of thought so that we can make design choices that our community of readers finds trustworthy and answer questions they find useful. Scholars attempting to describe and formalize these methodological traditions often present research processes and models of those processes. At the same time, it is common that these accounts reject rigid, linear models of the research process, even as they describe recurring patterns in how qualitative studies progress (Maxwell, 2012; Braun & Clarke, 2006).

Typically, the core critique of linear models is that qualitative research, even within a single discipline or paradigm, is not a fixed recipe. For example, Sarah Tracy conceptualizes qualitative research as *bricolage,* emphasizing that inquiry is necessarily adaptive, creative, and responsive to context rather than mechanical or strictly procedural (2010). Lucy Yardley similarly argues that *quality* in qualitative research comes not from the application of a set of standardized steps in a specific order, but from reflexive sensitivity to context, coherence, and impact of the work. *Interactive* models of qualitative design further reinforce this point by showing how goals, research questions, conceptual frameworks, methods, and validity considerations continually shape one another over the life of a study (Maxwell, 2012). Taken together, these foundational methods scholars emphasize to us that workflows in qualitative research are best understood not as linear pipelines but as flexible, evolving maps of analytic practice.

With awareness of these cautions, we nonetheless suggest that a qualitative research workflow has value for organizing our exploration of AI-assisted qualitative analysis. A workflow makes visible some of the key decision points in a research study that ultimately shape the final results. For example, decisions (including the order and number of decisions) on how to select cases, organize data, assign codes, aggregate themes, and create visualizations are all part of process between question and answer - and each decision will compress, reshape, or systematically remove aspects of an original qualitative phenomenon so that underlying patterns of interest may become more visible (Shaffer & Ruis, 2025). AI tools show promise for supporting many steps of many diverse human-led research processes. Organizing the most common and cross-cutting steps using a workflow helps organize our discussion of how AI is already used, where it may be helpful in future as the practice expands, and where it may threaten interpretive integrity.

For these reasons, the qualitative research workflow we propose in this chapter is intended as an organizing framework rather than a prescriptive method. It draws on foundational scholars across qualitative traditions such as grounded theory, thematic analysis, narrative inquiry, qualitative rigor and validity theory, and quantitative ethnography to identify a set of core stages that most qualitative studies traverse in some fashion. In the next section, we introduce this workflow and briefly describe each of its stages, setting the stage for the chapters that follow, which examine how these stages have been conducted historically, how AI is already reshaping them, and what is required to preserve rigor, trustworthiness, and reflexivity as new AI tools enter qualitative practice.

How You Should (and Should Not) Use this Workflow

Models of the qualitative research process are never neutral. Any workflow for qualitative research carries its own implications for how knowledge is produced, where authority resides, and what counts as rigor. Many qualitative traditions have long resisted the language of “pipelines” or standardized procedures because it can imply that the research process is uniform or mechanical, that there is an objectively “best” approach, or that the role of researcher is passive executor instead of active interpreter. This critique was foundational to qualitative inquiry since its inception, long before AI was available as a research tool.

With this framing, we present this workflow not as a universal or linear model for the qualitative research process, but instead as a shared reference point: a map, not a GPS navigator. A map does not generate a single linear route to a specific destination but shows you many possible ways to get to many valuable places. At each fork in the road, a map does not tell you which way to turn but makes all options visible and helps you assess the tradeoffs of each choice down the line. A map can help you communicate and justify your path to others, particularly to those who are lost or on a similar journey, so that you discover and build better routes together. Methods scholars have long pushed for this kind of process visibility, describing it in turns as auditability (Lincoln and Guba, 1985), transparency and sincerity (Tracy, 2024), or making the causal logic of a study explicit (Maxwell, 2012). Visibility is even more essential as we enter the era of AI-assisted qualitative research, where digital tools may obscure where and how interpretive transformation takes place unless there is very intentional human planning and monitoring.

Different qualitative traditions would, of course, engage with this workflow in different ways, and we celebrate this as an expression of the epistemic diversity in qualitative research. We encourage you to ignore, combine, reorder, or return to steps of this model to suit the goals, context, and priorities of your research communities and traditions, and to thoughtfully interrogate your use of artificial intelligence at every step. Our hope is that articulating a set of core stages that qualitative studies often traverse is a useful map: a common, if imperfect structure for thinking about common entry-points for AI support while retaining a shared language for ongoing negotiations of rigor, reflexivity, and the consequences of our analytic choices.

A Qualitative Research Workflow for AI Inclusion

To organize the discussions that follow, we propose a qualitative research workflow that identifies a set of analytic stages that are often incorporated in qualitative research. These stages are highlighted as common points where methodological decisions are made and are possible entry-points for the incorporation of artificial intelligence. Again, our ordering below does not imply linearity or a required order, though many of these steps *will* occur in this order, in many projects.

The upcoming chapters of this book are organized around eleven stages of qualitative research, shown in Figure 5-1:

1. **Framing a Qualitative Research Study**
   Defining the phenomenon, theoretical lenses, paradigmatic commitments, and research questions that guide qualitative inquiry.

2. **Collecting and Selecting Data**
   Making decisions about sampling, cases, data sources, and inclusion that determine what materials are created, collected and examined.

3. **Preparing Data**
   Systematically transforming materials into analyzable forms through cleaning, transcription, formatting, and documentation of information loss.

4. **Immersion, Familiarization, and Early Sensemaking**
   Engaging deeply with the data through reading, memoing, and reflexive attention to what is present and what is missing in the data.

5. **Segmenting and Structuring Data: Unitizing Meaning**
   Defining analytic units, segmenting texts or events, and creating the structures that make analysis possible.

6. **Developing Codes and Codebooks**
   Creating analytic categories, whether theory-driven, data-driven, or hybrid, that specify what kinds of meaning are being tracked.

7. **Refining Codes Over Time**
   Iteratively revising codes through saturation, comparison, and theoretical development as understanding deepens.

8. **Applying Codes**
   Assigning codes to data using human judgment, AI assistance, or automated systems at varying scales.

9. **Validation, Trustworthiness, and Quality Checks**
   Evaluating the credibility, coherence, and accountability of findings using paradigm-appropriate practices.

10. **Interpretation, Sensemaking, and Theory Building**
    Integrating coded data into higher-level explanations, narratives, or theoretical models.

11. **Analysis of Coded Data**
    Examining patterns, relationships, and structures across cases using qualitative, mixed-methods, or computational analytic techniques.

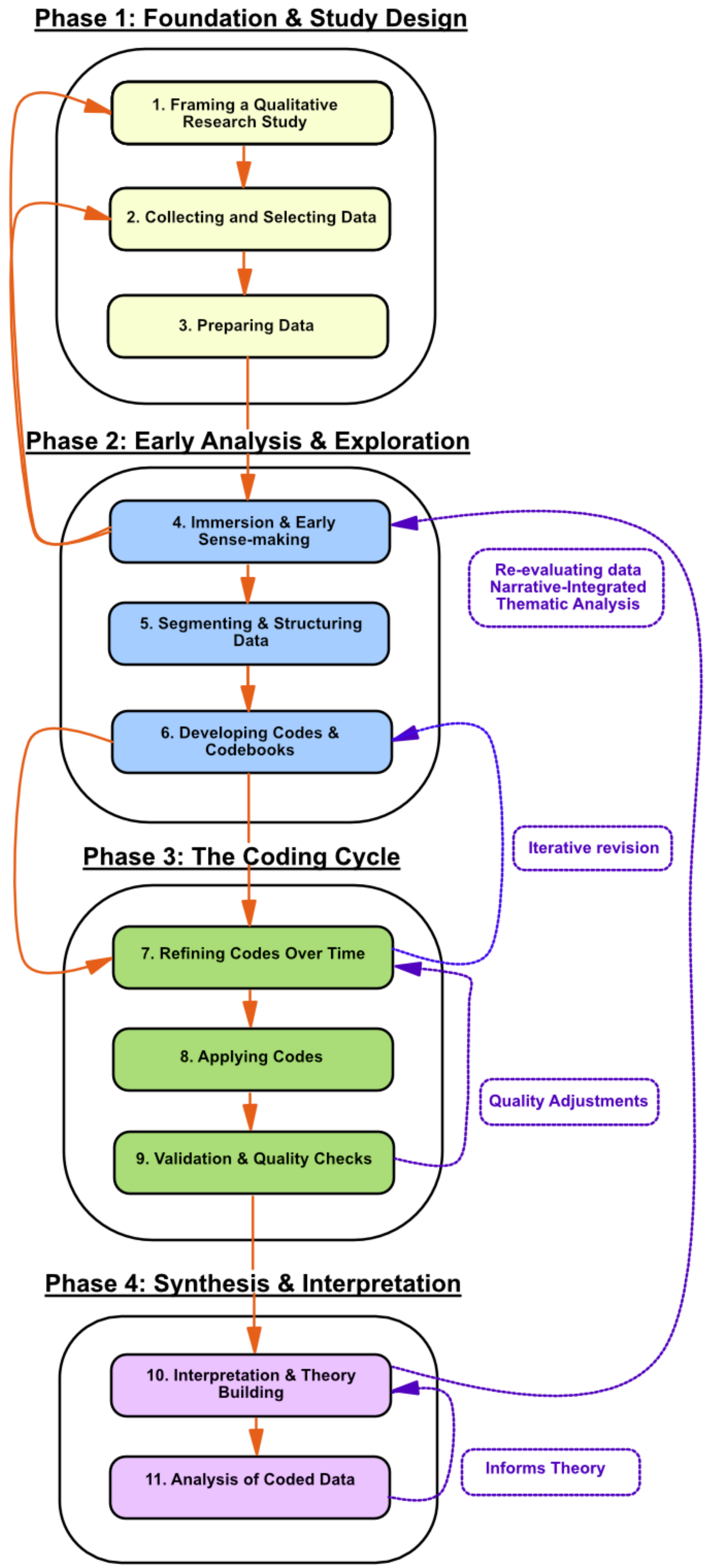
Phase 1: Foundation & Study Design
1. Framing a Qualitative Research Study
2. Collecting and Selecting Data
3. Preparing Data
Phase 2: Early Analysis & Exploration
4. Immersion & Early Sense-making
5. Segmenting & Structuring Data
6. Developing Codes & Codebooks
Re-evaluating data Narrative-Integrated Thematic Analysis
Phase 3: The Coding Cycle
Iterative revision
7. Refining Codes Over Time
8. Applying Codes
9. Validation & Quality Checks
Quality Adjustments
Phase 4: Synthesis & Interpretation
10. Interpretation & Theory Building
11. Analysis of Coded Data
Informs Theory

Figure 5-1. The Qualitative Research Process used within this book

In the sections that follow, we briefly introduce each stage, describe what it typically involves in qualitative research, and highlight key ways that artificial intelligence may play a current or future role. We will return to each of these stages in deeper detail, within their respective chapters.

(1) Framing a Qualitative Research Study

Framing a qualitative research study involves articulating what a prospective qualitative study is *about*, such as defining the scope, purpose, and research context. This often involves the preparation of a research question (though sometimes these are changed or iterated over time), the grounding of both the researchers and their study design in specific ways of thinking (ontology, epistemology) that will prioritize or de-prioritize certain research steps, making preliminary connections to any relevant bodies of existing research on the topic, and suggestions about the kinds of emphasis, language, and methodologies might be best suited to enact the work. Put simply, framing a qualitative research study often includes deciding what to study (the problem and context), why it matters (purpose), how to best approach it (methodology), and who the audience is.

Across different qualitative traditions, study framing is often described as the foundation of a research inquiry: Braun and Clarke (2013) emphasize the importance of clarifying analytic aims and theoretical positioning before and during analysis, Creswell and Creswell (2018) introduce research questions and research paradigms as the organizing backbone of a study's design, and Ravitch and Carl (2019) highlight the value of a conceptual framework as the connection between theory, purpose, and method. Even in approaches that favor the open-ended exploration of a phenomenon, such as grounded theory (which de-emphasizes aspects of pre-study preparation so that meaning can emerge; Bryant & Charmaz, 2011), early framing decisions are still important to shape what is examined, what is noticed, and what counts as meaningful.

It's important to of course note that framing choices are not limited to the initial stages of qualitative work but may be (and typically are) routinely revisited as a study evolves. Research questions may be refined or new questions asked as insights are uncovered, theoretical commitments may be sharpened or shifted, and the practices being used are adjusted as researchers engage with and learn from their phenomena of focus. For this reason, Maxwell (2012) characterizes the entire life cycle of a qualitative research study as an interactive process in which goals, conceptual frameworks, research questions, and methods continually inform one another rather than proceeding in a fixed sequence. From this perspective, a qualitative inquiry is *always* being framed and reframed. In this sense, framing a qualitative research study should not be understood as a bounded stage at the beginning of a project, but an ongoing act of sensemaking that anchors the entire process.

For reasons we have discussed in prior chapters, the intentional framing of a study becomes especially consequential when considering the inclusion of AI. Core decisions around what questions are worth asking, what ontologies or epistemologies are most appropriate, or what analytical actions and decisions are optimal for your context, ultimately cannot be offloaded to an AI. To mis-quote Sartre (1946), “I can always choose, but I must know that if I let AI choose for me, that is still a choice.” With thoughtful training and/or prompting, AI systems may operate within the boundaries set by human researchers’ (though probably also their designers’) conceptual and theoretical choices. Ultimately, if these framing commitments are left implicit or are poorly articulated, AI tools may produce insights that appear to align with their user’s framing, but in practice may reflect other implicit assumptions and foundations. Chapter #’s discussion of Framing a Qualitative Research Study explores how AI may support study framing without usurping the human researcher’s necessary role, and how study framing may be introduced to an AI tool to better align its output in later research steps.

(2) Collecting and Selecting Data

In qualitative research, data collection and data selection (sometimes referred to as data sampling) typically involves a set of decisions about what and how empirical materials will be gathered and used as representations of a phenomenon of interest. Depending on research goals and contexts, this may include (1) identifying cases, sites, participants, and artifacts, (2) establishing criteria for inclusion and exclusion in a study, and (3) deciding which data representations will be created, collected, and used (e.g., interviews, observations, documents, digital traces). Across qualitative traditions, these choices are often pragmatic (like when a researcher must use a convenience sample of only willing participants), but are never exclusively logistical; the framing of a study’s goals, guiding frameworks, and research questions should also inform what data sources are the best fit to help to answer the research question (Maxwell, 2012; Ravitch and Carl, 2021; etc.). In many traditions, data collection is also cyclical, with early rounds of data analysis informing what data researchers need to collect next, and sampling shifting as new insights emerge.

Some qualitative traditions are framed around the idea that data is not passively “collected” but is instead actively “produced” through research relationships and practices. Grounded theory (Bryant & Charmaz, 2010), narrative inquiry (Riessman, 2008), and to an extent discourse analysis (Gee, 2025) all organize around this idea: that researchers engage with participants, texts, and settings, and this interaction shapes what the data ultimately is. Within these traditions, across the life cycle of a qualitative study, interview questions, observation protocols, transcription practices, and even the social positioning of the researcher will shape (at least in part) what is said, documented, and made available for analysis. Therefore, data collection and interpretation are necessarily intertwined. As such, data collection and selection can appear as a more or less discrete step in the process of qualitative research, with conceptualizations, iterations, and best practices varying by research tradition.

AI tools can support data collection and selection in a variety of ways; the potential uses vary by research tradition. They can equip researchers to expand the scale and scope of the data that is included in a study, from informing procedures and providing programming language needed to access large digital data archives, to summarizing structural aspects of a dataset, to supporting case selection, to even assisting with interviewing. Chapter # will discuss the ways that AI can play a role in data collection and selection while remaining aligned with your study's goals framing, and ethical commitments.

(3) Preparing Data

Data preparation involves the transformation of raw materials you choose to include or generate so that they are better suited for examination (a phase often called data analysis). This can include activities that change the data's format, such as audio transcription or the reorganization of text. It can also involve the systematic removal of data, such as anonymization (removal of identifying information). These steps are sometimes overlooked in qualitative methods texts because they tend to be more technical, pragmatic, and idiosyncratic to a specific study, but a growing body of methodological work emphasizes that data preparation is itself a deeply interpretive and consequential step in qualitative research (Braun & Clarke, 2013; Miles et al., 2014; Zörgő et al., 2021). Discourse analysts specifically have long argued that decisions about what data to include, what to omit, and how to organize speech or action ultimately impact the outcomes of a research study (Gee, 1993). From this perspective, preparing data is not just about making it usable, but about intentionally and practically constructing the analytic object.

Quantitative ethnography (QE) and related mixed-methods traditions make this point especially explicit. In QE, data is structured through systematic, context-driven rules chosen by the researcher so that patterns in the data may be understood through visualization and comparison. These transformations and distillations are a valuable part of the process: data "noise" that is not relevant to the question of inquiry are filtered out, while patterns of complexity that are important to the researcher are made more visible. In addition to choosing what data preparation processes are needed, researchers at this step also choose what approaches they use to document, justify, or test the appropriateness of their distillation choices. Again, we agree with Sartre (1946): choosing not to do so is itself a choice.

AI tools show promise but also risks at the data preparation stage. Tools for automated transcription, translation, cleaning, summarization, and feature extraction have emerged and improved in quality over time (e.g., Radford et al., 2023), facilitating analysis as data sources grow in complexity, accessibility, and scale. On the other hand, errors in all these tasks remain common even for state-of-the-art technology and can exacerbate the already lossy nature of data transformations. We dedicate chapter # to a discussion of the ways AI can support data preparation processes, including the many affordances, current best practices, and common (and significant) pitfalls.

(4) Immersion, Familiarization, and Early Sensemaking

While some more structured models of the qualitative research process (Creswell & Creswell, 2018; Miles et al., 2014) recommend a transition to data analysis after data collection and preparation, the process of data "analysis" itself actually encompasses a variety of goals and steps that we parse with more specificity here. This chapter highlights an important preparatory step that often takes place in earlier phases of qualitative research: the iterative process of immersion, familiarization, and early sensemaking of data. In qualitative research texts, these terms describe one or more periods of sustained, attentive engagement with a dataset (sometimes described as "adopting an attitude of noticing"; Seidel, 1998) during which researchers can begin immersing themselves in the structure and context of the data. In contrast to later analytical reviews of qualitative data with a study-specific purpose (e.g., finding data-driven themes, coding), this step involves repeated reading, viewing, or listening to data and noting initial patterns, surprises, and tensions that emerge.

Although this stage is sometimes treated as part of data collection or preparation, separating it conceptually highlights a shift in the researcher's stance, important to some qualitative paradigms. While the collection and preparation of data establish what materials exist and the form in which they might be useful, immersion and familiarization involve beginning to think with those materials. This is the point at which researchers start to notice recurring ideas, striking moments, absences, and contradictions, and to reflect on how their own assumptions, theoretical commitments, and positionalities shape what stands out. Maxwell (2013) frames this as the beginning of analytic thinking, where tentative explanations and hunches arise that will later be tested and refined. Bryant and Charmaz (2010) calls early engagement with data an interpretive act for researchers, who are actively constructing meanings rather than simply uncovering them.

These processes often align with anthropologic and ethnographic traditions, where researchers embedded themselves in the complexities of a phenomenon for long periods to develop immersive understandings. At this stage, scholars may also work to develop a reflexive orientation to both the material and their own responses to it by writing or affirming their positionalities or writing analytic memos. Braun and Clarke's qualitative research guidelines call this phase "familiarization" as a foundational step in thematic analysis, during which researchers actively read and reread their data to become deeply acquainted with its breadth and depth. Saldaña highlights the creation of researcher memos as a way to externalize emerging insights and questions before formal coding begins. In grounded theory and other inductive approaches, early sensemaking is often tightly intertwined with ongoing data collection, but it nonetheless represents a distinct analytic activity: the work of orienting oneself to what the data are saying.

AI introduces both opportunities and risks at this stage. Tools that generate summaries, topic models, keyword lists, or semantic embeddings can provide rapid overviews of large datasets and help researchers identify areas of interest that might otherwise be difficult to locate. However, if these tools replace rather than support this human immersion, they can short-circuit the slow, reflexive engagement that underpins qualitative understanding (a concern raised in Jowsey et al., 2025). Early sensemaking is not only about detecting patterns, but also about developing a situated feel for the data that informs later analytic choices. Treating immersion and familiarization as a distinct stage of the workflow helps clarify where AI can assist by expanding researchers' reach, such as in the cases where datasets are too large or complex to engage with in totality, or where it might be able to draw on alternate frames of understanding that serve as useful counterpoints to the researchers' "noticings" and reflexive perspectives.

(5) Segmenting and Structuring Data: Unitizing Meaning

Segmenting and structuring data is the stage in which researchers decide how qualitative material will be grouped and organized into analytic "pieces." This process can be explicit to varying degrees between qualitative traditions. For example, in quantitative ethnography (Shaffer & Ruis, 2025), this includes segmenting the dataset into "lines", treated as the most refined pieces of data that retain interpretive value for the research inquiry (e.g., a turn of talk, a sentence, an episode, a document, a time window, even a single word). These lines are then organized into "units" which are groups that represent structure in the data that can help answer specific questions (e.g., grouping lines of interviews by participant, grouping lines of classroom data by day, etc.). Metadata and indexing structures are typically required to allow those units to be tracked, compared, and revisited. While this step of data structuring is left implicit in many models of qualitative analysis, choices at this step are impactful. Every act of coding, interpretation, or pattern detection is built upon the prior decision about what is being examined and how it should be organized.

Different qualitative traditions approach unitizing in different ways. In discourse analysis and narrative inquiry, scholars such as Gee (2025) and Riessman (2008) emphasize that how texts are segmented—into clauses, turns, narratives, or discursive moves—shapes what kinds of meanings can be analyzed. In thematic and grounded theory approaches, Saldaña (2011) and Bryant and Charmaz (2011) describe coding units that may range from single words to extended passages, depending on analytic purpose. Ultimately, the act of segmentation is not neutral, but instead reflects theoretical assumptions about what constitutes an action, a meaning, or a phenomenon, and governs what can later be constructed or gathered.

AI systems are particularly sensitive to the choices made at this stage. Most computational methods operate over discrete units (e.g., chunks of text, tokens, utterances, time windows) and the way data are segmented will strongly influence what patterns those systems can detect. An AI tool instructed to analyze an interview as a whole narrative, as a set of individual utterances, or as a collection of overlapping thematic segments will highlight different kinds of structure. Automated segmentation and chunking tools can help scale qualitative analysis to large corpora, but they also risk imposing boundaries that do not align with researchers' theoretical or interpretive aims. Treating unitizing and structuring as a distinct stage of the workflow makes these boundary-setting decisions visible and open to critique, ensuring that AI-assisted analyses remain grounded in conceptually meaningful representations of qualitative data at hand, and not arbitrary or purely technical partitions.

(6) Developing Codes and Codebooks

Developing more formalized codes and codebooks is a common yet highly diverse stage in recent qualitative tradition, during which researchers define the analytic categories that may be later used to label, organize, and interpret segments of data. Codes specify what kinds of meanings, actions, or phenomena the study is attending to. A codebook articulates and formalizes how those codes are defined, distinguished, and applied. Across qualitative traditions, this step is where theoretical commitments, empirical observations, and analytic purposes converge. Saldaña (2016) describes coding as the process of linking data to concepts through iterative cycles of labeling and refinement. In grounded theory, (e.g., Corbin & Strauss, 2014) codes are seen as emerging through constant comparison, as researchers move back and forth between data and developing ideas. In more theory-driven traditions, codes may be derived from prior literature, conceptual frameworks, or analytic lenses specified during the framing stage (e.g. Miles et al., 2014; Boyatzis, 1998). Many individual studies adopt hybrid approaches, in which initial deductive codes are revised, expanded, or replaced in response to patterns discovered in the data.

Like many steps mentioned prior, the choices made during code development are foundational for everything that follows. Codes determine what will be made visible in the data and what will remain in the background, which in turn structures subsequent analyses, displays, and interpretations. Bryant and Charmaz (2011) stress that codes are not neutral descriptors but analytic constructions that reflect the researcher's goals, perspectives, and theoretical interests. Even more strongly, Braun and Clarke (2006) argue that coding is inseparable from meaning-making, because it actively produces the themes and narratives that will later be developed. Whether codes are broad or fine-grained, stable or provisional, theory-laden or open-ended, shapes the kinds of claims a study can ultimately support.

AI systems are increasingly being used to suggest codes, making this stage a central site of human–AI interaction. Large language models can propose candidate codes, cluster similar cases, or map data-driven constructs to predefined theories at a scale that would be time-consuming for human coders alone. Early research suggests that AI tools can help expand scholars' thinking during code development and ultimately result in codebooks that are more exhaustive (Barany et al., 2024, Bijker et al., 2024; Wen et al., 2025). New work in this space has also begun to question whether formal coding frameworks are even a necessary step when working with LLMs, proposing alternative approaches that emphasize dialogic querying, narrative synthesis, or theme generation without a predefined codebook (e.g., Friese, 2026; Nguyen-Trung & Nguyen, 2025). However, without careful and consistent human oversight, these tools can and will reify superficial patterns, reproduce biases embedded in training data, miss categories that human analysts would identify, or collapse theoretically important distinctions. While treating code development as a step where AI can support researchers, we nonetheless remind the reader that researchers remain responsible for deciding what counts as a meaningful code, how it is defined, and how it aligns with the study's conceptual and epistemic commitments.

(7) Refining Codes Over Time

Refining codes over time is the stage (and we use the term stage here especially loosely) in which analytic categories are tested, revised, split, merged, and sometimes abandoned as researchers learn more about their data and their emerging interpretations. While code development establishes an initial analytic lens, refinement is the process through which that lens is sharpened. In grounded theory, this work is often described in terms of constant comparison, theoretical saturation, and the movement from provisional to more conceptually robust categories. Saldaña (2011) emphasizes multiple coding cycles, in which early labels give way to more focused, patterned, or theoretically informed codes. Braun and Clarke (2013) note that themes are not simply discovered but actively developed through iterative engagement with coded data, as researchers revisit earlier decisions in light of later insights.

Perhaps more than any other step, the code refinement stage is rarely linear or confined to a single moment in the research process. Codes routinely evolve in qualitative research as new data are collected, as researchers encounter disconfirming cases, or as analytic priorities shift. Maxwell (2013) frames this as part of an interactive model in which theory, data, and interpretation continually shape one another. Bryant and Charmaz (2011) stresses that analytic categories remain provisional, grounded in both empirical comparison and the researcher's evolving conceptual understanding. Whether researchers speak in terms of saturation, stability, or coherence, the underlying idea is the same: categories must be sufficiently well-developed to support meaningful interpretation, but never so rigid that they resist changes to better suit the needs of a project or context.

AI systems have potential to both complicate and potentially accelerate code refinement. Automated coding, clustering, and similarity detection can reveal inconsistencies, overlaps, or drift in code usage that might be difficult for humans to detect at scale. They can also suggest strategies for refining codes if given proper context, training and prompting. At the same time, AI systems trained on fixed definitions can prematurely stabilize categories, making them appear more settled or prevalent than the underlying theory or data warrants. Treating code refinement as a distinct yet ongoing stage of the workflow may help guard against this risk by emphasizing that analytic categories are always subject to revision. Whether codes are applied by humans, machines, or both, the responsibility for deciding when a construct is sufficiently saturated, coherent, or theoretically meaningful remains fundamentally in the hands of the human analyst.

(8) Applying Codes

Code application, or "coding", is the stage in which analytic categories (codes and codebooks) may be used to systematically label, organize, and annotate segments of qualitative data. In this stage, theoretical and conceptual decisions made during code development and refinement are operationalized across a dataset. In other words, code development and refinement tell a researcher what relevant themes might be present in a dataset, and coding reveals whether, where, and how much the code is present in the data. Coding can also give numeric structure to qualitative data (e.g., binary coding) if the researcher plans to visualize, compare, or numerical summarize patterns of coding in later analysis. Saldaña (2011) emphasizes this process as a deeply transformational act because each coding decision is an individual judgment about what a piece of data means and how it relates to the study's evolving concepts.

Different traditions and project contexts place different demands on this stage. In small-scale interpretive or exploratory studies, coding may be tightly interwoven with memoing and interpretation, with researchers cycling fluidly between labeling, reflecting, and theorizing. In larger or team-based projects, formal codebooks, training, and consistency checks may be introduced to support shared understanding and comparability. Quantitative ethnography and mixed methods approaches further treat coded data as inputs to subsequent modeling or network analysis, making the reliability and traceability of coding decisions especially consequential. Across these contexts, key decisions include who codes, how much training or calibration is required, and how disagreements or ambiguities are handled.

Due to their complex language-processing capabilities, large language models have the potential to support this stage. With training, machine-assisted and automated coding systems can apply codes across large datasets quickly and comparably as consistently as human decision-making, making it possible to work at scales that would otherwise be infeasible. Yet this efficiency comes with trade-offs: like human coders, these systems can assign a concrete label that masks conceptual ambiguity, and their outputs may involve or embed theoretical drift. Attention must therefore be paid to interpreting edge cases and ensuring that automated outputs align with the study's analytic intent. Ultimately, AI can change how coding is done, but not what it means -- coding remains an interpretive act, even when some of the labor is delegated to machines.

(9) Validation, Trustworthiness, and Quality Checks

Validation and quality assurance are the stages in which researchers examine whether their analyses are credible, coherent, and warranted given their data and theoretical commitments. Like every step in the research process, validation is not a single definition or set of measurements but is shaped and defined by multiple paradigm-aligned criteria such as credibility, plausibility, rigor, reflexivity, and visibility. Lincoln (1985) frames this stage in terms of trustworthiness (e.g., credibility, dependability, confirmability, transferability), while Maxwell (2012) emphasizes validity as the quality of inferences that researchers draw from their data. Tracy (2010) and Yardley (2000) similarly argue for sincerity, resonance, transparency, and coherence, a set of flexible but principled criteria that can help readers evaluate the integrity of qualitative claims.

This stage can include a range of analytic practices: member checking, peer debriefing, audit trails, triangulation, reflexive memoing, intercoder agreement, negative case analysis, and thick description, among others. Which practices are appropriate depends on the study's epistemological stance and analytic goals. Grounded theorists may prioritize theoretical saturation and constant comparison (which happens throughout a study), while narrative or discourse analysts may focus on interpretive plausibility and contextual richness. Quantitative ethnography and mixed-methods traditions often add formal reliability or model-fit checks to ensure that coded and structured data support valid downstream analyses. Across all approaches, the core aim is the same: to ensure that analytic claims are well grounded, transparent, and accountable.

AI raises both opportunities and risks for validation. Automated systems can make coding and analysis more consistent and auditable, generating logs and reproducible outputs. If prompted, they can share rationales for their decisions that can be used to address inconsistencies. At the same time, scholars have raised the concern that these outputs may create a false sense of transparency and trustworthiness, given that LLM rationales may not reflect actual reasoning (Bender et al., 2021; van Dis et al., 2023).

(10) Analysis of Coded Data

In the doctoral seminar she teaches, the second author often describes segmenting data (See Chapter #) and applying codes (See Chapter #) as types of data "systematization" because these processes impose structure. For example, a paragraph may be segmented into sentences (one sentence = one "unit" of analysis), and each unit may be binary coded for the presence (1) or absence (0) of a code. Analysis of coded data is the stage in which researchers may use data that has been "systematized" to conduct additional analysis that can help reveal useful patterns or connections. This can be something as simple as descriptive statistics, or can include structured modeling, visualization, clustering or even predictive techniques. These supplementary analyses can help to answer research questions such as "which themes tend to co-occur?" "How do patterns differ across participants, settings, or conditions?" or "What trajectories or configurations emerge over time?" This work may be also carried out through matrices, displays, comparisons, and narrative syntheses (e.g., Miles et al., 2014).

In quantitative ethnography and related approaches, this stage includes formal modeling and visualization techniques such as epistemic network analysis (Shaffer et al., 2016), clustering, sequence analysis, or other pattern-detection methods. These techniques allow researchers to explore the structure of coded data at scale, identifying regularities, contrasts, and emergent groupings that would be difficult to see through close reading alone. Researchers in other traditions such as interpretive phenomenological analysis or narrative inquiry might instead return to the coded data to construct detailed case comparisons, trace narrative arcs across participants, or develop thematic matrices that preserve the richness of individual accounts while enabling systematic cross-case interpretation.

AI expands what is possible at this stage by enabling large-scale pattern detection, automated comparisons, and visualizations over massive corpora. AI tools may also be able to offer suggestions for which analytical processes are most appropriate for a data inquiry and context and offer assistance with crafting programming language for completing analyses. Overall, there is growing potential for AI to provide new lenses through which coded qualitative material can be explored, compared, and ultimately brought back into human sensemaking.

(11) Interpretation, Later Sensemaking, and Theory Building

Interpretation and sensemaking are the stages in which researchers may start to move from coded and organized data to a final accumulation of analytic claims, narratives, and theoretical contributions. This is where patterns are explained, meanings are articulated, and findings are woven into accounts that speak to the study's research questions, conceptual frameworks, and broader scholarly conversations. Analytic writing, memoing, and theorizing are again central here. Riessman (2008) and Gee (2025) characterize this step as the construction of analytic stories about what the data show and why it matters. Though we present this as a stage for comprehensibility, Maxwell, Charmaz, and Tracy all emphasize that interpretation is not something that happens after analysis is "complete," but is produced through the act of writing, comparing, and reflecting on data.

This stage includes decisions about what counts as a finding, how evidence is organized to support claims, and how results are connected to prior research and, where applicable, theory. Some traditions aim to generate grounded theory, others to produce thick description, critical interpretation, or narrative accounts. Issues such as generalizability, transferability, and theoretical contribution are addressed in this stage, in paradigms emphasizing these values. However, rather than asking whether results are generalizable, many qualitative scholars instead frame this stage as asking how findings illuminate processes, meanings, or mechanisms may resonate beyond the specific cases studied. In quantitative ethnography and related mixed methods approaches, this stage also involves what these communities call "closing the interpretive loop" by connecting statistical or visual patterns back to the underlying qualitative data. Models and metrics are treated not as endpoints, but as prompts for renewed sensemaking, requiring researchers to revisit not only the data, but also every prior analytic decision that helped to generate them. This preserves the core qualitative commitment that meaning resides in the data and its interpretation, and not in numerical abstraction.

AI may support this stage by helping researchers surface patterns, compare cases, summarize coded data, or explore alternative interpretations. During this stage, questions such as what a pattern means, which interpretations are theoretically significant, or how findings should be framed for a given audience continue to find their expression in terms of the researcher's theoretical commitments, positionality, and scholarly aims. While AI may assist with organizing and exploring analytic material, the work of meaning-making still remains in the mind that makes that meaning.

Variation Across Traditions and Disciplines

Although we present these eleven stages as a shared qualitative research workflow, different traditions and disciplines emphasize, reorder, or even bypass particular steps. Grounded theory foregrounds iterative data collection and code refinement. Narrative inquiry traditions emphasize steps related to data interpretation and representation. Quantitative ethnography formalizes segmentation and structuring of data and heavily adopts post-coding analyses. Applied or computational mixed methods approaches may prioritize scaling, automation, or statistical pattern detection. None of these approaches follows the workflow in a single, uniform way. We include this workflow not to prescribe how qualitative research ought to be done, but to make visible the set of analytic moves that different traditions make in different ways. This matters especially for AI-supported research, because the opportunities and risks of automation, modeling, or augmentation depend on which stages are emphasized and how they are enacted. Our goal in this chapter is therefore to offer a common language for thinking about AI practices across fields while still preserving the methodological pluralism that defines qualitative inquiry.

How this Book is Organized

The remainder of this book is organized around the eleven stages of the qualitative research workflow described above. Each subsequent chapter focuses on one stage of the process, using it as a lens for examining how AI is reshaping and can reshape qualitative practice. Within each chapter, we first describe what the stage typically involves and how it has been carried out in pre-AI qualitative research traditions. We then examine where AI tools are already being used, where they could plausibly be integrated in the future, and what kinds of human oversight and methodological care are required to maintain rigor, trustworthiness, and reflexivity.

Each chapter also highlights key concerns, tradeoffs, and challenges associated with introducing AI at that point in the workflow, with attention to what impacts might emerge across the lifespan of the study. We pair these discussions with case studies drawn from existing research, illustrating how particular design choices, tools, and analytic strategies play out in practice. Together, these cases ground the abstract workflow in concrete scholarly work across disciplines, methods, and scales.

By structuring the book around a shared qualitative workflow, we aim to provide readers with a practical map for navigating human–AI partnerships in qualitative research. The chapters that follow do not prescribe a single “right” way to use AI but instead offer a set of analytically grounded entry points for making intentional, transparent, and theoretically informed decisions. We hope to support a future in which AI expands what qualitative researchers can do, enhancing and not eroding the interpretive, ethical, and epistemic commitments that underpin qualitative research.

Find the most up-to-date version of this chapter’s citations at
https://learninganalytics.upenn.edu/book/ai-qualitative-research.html