International Research in Education
ISSN 2327-5499
2026, Vol. 14, No. 1

Rigor, Trustworthiness, and Credibility in Qualitative Research: The Qualitative Forensic Audit Trail Tool—Artificial Intelligence (QFATT-AI)

David C. Coker

Associate Professor and Doctoral Methodologist

College of Education and Human Performance

West Liberty University, United States

Received: December 18, 2025 Accepted: January 5, 2026 Published: January 7, 2026

doi: 10.5296/ire.v14i1.23507 URL: https://doi.org/10.5296/ire.v14i1.23507

Abstract

Qualitative research grapples with persistent questions of trustworthiness and rigor, often lacking the definitive validation metrics of quantitative studies. The audit trail, a systematic documentation of research decisions and methodologies, was proposed in the 1980s as a solution to enhance trustworthiness and credibility. Contemporary implementation has drifted significantly from its original intent, often reduced to a perfunctory checklist or post hoc justification rather than a rigorous analytical tool. This article critiques the current state of audit trails, revealing a conceptual dilution where audit trails serve as mere stamps of approval, devoid of critical engagement or documented discrepancies. In response, this paper introduces the Qualitative Forensic Audit Trail Tool—Artificial Intelligence (QFATT-AI), a novel framework that reimagines the audit trail for the digital age. By integrating generative AI as an external, dialectical partner, the QFATT-AI provides a structured method for challenging researcher bias, improving methodological fidelity, and exploring alternative interpretations through upside-down thinking. We propose the Transparent Reporting of AI Logistics (TRAIL) to improve ethical and clear documentation of AI involvement. This approach moves beyond simple verification to a robust process of falsification and plausibility, offering researchers a practical means to reclaim the audit trail as a dynamic engine for quality and credibility in qualitative inquiry. The QFATT-AI represents an evolution in qualitative research, transforming passive documentation into active, rigorous interrogation.

Keywords: audit trail, trustworthiness and credibility, qualitative research, research methodologies, artificial intelligence

1. Introduction

"An audit trail not only enhances the reliability of the study by making it open to scrutiny but also contributes to the field's collective knowledge by offering a roadmap for future research," (Ranaweera, 2024, p. 324). Weil (2008) inquired whether a legitimation crisis existed, suggesting that methods such as triangulation, audit trails, and reflexivity could help address doubts about rigor and reliability. Qualitative research is inherently subjective and lacks the positivist assumptions of quantitative research, requiring researchers to justify claims and present evidence for rejected possibilities (Wesley, 2014). Many techniques exist to improve trustworthiness and credibility, but a general lack of methods or erudition often renders many claims mere assertions, with the absence of divergence or the need for improvement never addressed (Coker, 2023). A way to improve reliability and validity, suggested long ago, was to develop an audit trail that used the complete research process to analyze and improve results by making processes transparent and checking research bias (Fetterman, 1988; Greene et al., 1988). Trustworthiness and credibility in qualitative research depend on systematic, transparent methods that document prior practices and improve results.

Halpern proposed that qualitative research could be improved by using an audit trail, which involves documenting the processes by which practices are refined during a study, examining decision-making post hoc, or increasing transparency to make the process more transparent (Leko et al., 2021; Nowell et al., 2017). For many, audit trails quickly transformed into post hoc endeavors, with questionable evidence of effectiveness (Krefting, 1991). Audit trails can be a complementary vehicle for data analysis, but often the researcher's thinking is not well documented or knowable (Cutcliffe & McKenna, 1999). Akkerman et al. (2008) proposed three standards: visibility, comprehensibility, and acceptability by examining if the results were grounded in the data and reasonable interpretations. What audit trails cannot do is turn poor methods and instrumentation into a rigorous, thorough study (Korstjens & Moser, 2017).

Audit trails are popular in research for assessing confirmability and dependability, but a gap exists in the practices between what is claimed versus the procedures, with audit trails often little more than a means to claim trustworthiness and reliability, with most claiming no discrepancies or need for improvements (Coker, 2023; Megheirkouni & Moir, 2023). Publications often limit article length, obscuring what auditors do and preventing readers from evaluating claims (De Kleijn & Van Leeuwen, 2018). Cutcliffe and McKenna (2004) pointed out the need to think outside the box, which guides the inquiry in this article. First, a literature review examines the origins and development of the audit trail since the 1980s. Secondly, qualitative articles from the past 15 years are collated and analyzed thematically to identify trends. Finally, following Cutcliffe and McKenna's call years ago, a bold reimagining using artificial intelligence (AI) provides an ethical framework, a systematic application of AI to all qualitative research, and the use of the qualitative forensic audit trail tool.

2. Background

In 1985, Lincoln and Guba published the seminal *Naturalistic Inquiry*, redefining rigor and relevance from a quantitative, positivist framework of reliability and validity to the concepts of dependability, confirmability, credibility, and transferability, popularly named under the moniker trustworthiness and credibility (Lincoln & Guba, 1985). Part of Guba's world was the young student Halpern, who, under his tutelage, proposed developing an audit trail —a concept borrowed from accounting, where an external expert in the methodology and field reviews for discrepancies and examines whether receipts—records—exist to support the analysis and results. "Although reliability and validity are treated separately in quantitative studies, these terms are not viewed separately in qualitative research" (Golafshani, 2003, p. 600). Though Owens (1982) proposed developing an audit trail more than a log, Halpern's (1983) dissertation proposed three central aims for using audit trails to improve the transparency, rigor, and relevance of qualitative research: fidelity of methods with implementation, rigor and adequacy of results from data collection, and minimization of bias in analysis. The purpose of the background is to report the history of the audit trail since its inception in the 1980s.

An audit trail is a systematic method for measuring the four components identified by Lincoln and Guba, providing a complete accounting of all materials used in planning, executing, and reporting a study, and cataloguing and sorting data before a qualitative study for use in an audit (Gerson, 1984). An audit trail, sometimes called a decision trail, transparently reports to readers for adjudication that results logically connect to the research plan and data, and provides a window into the researcher's decisions (Bowen, 2009). The systematic application of an audit trail enhances trustworthiness and credibility, supports the reporting of results that are both formative and summative, is grounded in the data in a logical and coherent manner, and justifies the iterative and recursive nature of the research (Greene et al., 1988; Miller, 1997). A well-executed audit trail displays the chain of evidence and how researchers linked data to inform decisions that led to the results.

Credibility in an audit trail is a foundational principle of internal validity that asserts that results describe a phenomenon as lived by participants, are transparently reported, demonstrate deductive soundness, are supported by empirical data, and offer plausibility over probability by examining competing arguments (Korstjens & Moser, 2017; Wesley, 2014). An example is Braun and Clarke's (2006) clear admonition in thematic analysis that researchers consider candidate themes, or themes that offer alternative possibilities to challenge researcher biases or foreclose results too early. Confirmability audits examine the nature of inquiry and decision-making to form interpretations that the data warrant the results (Akkerman et al., 2008), but Coker (2026) points out that explication beyond what emerges often defies the records and requires a researcher to document personal thinking in vivo, which often does not happen. Qualitative research walks a fine line between opinion and research, and an audit trail is a way to make transparent the analytical decisions that yield results beyond anecdotalism or cherry-picked quotes.

Researchers must determine the objective of audits, which may be formative or summative, used during the research process, or for post hoc analysis. Another primary concern is that audit trails can be used to assess the sufficiency and quality of interpretations or as an analytical tool to refine and revise proposed results (De Kleijn & Van Leeuwen, 2018). A key purpose of an audit trail is not to generate new results or replicate prior findings, but to evaluate whether researchers' actions and findings are supported by the methods, data collection, and data analysis (Greene et al., 1988). Halpern (1983) provided examples in his appendices, in which an auditor sampled data collection to evaluate whether the results could be backward-engineered to be supported by the research process. The purpose was not necessarily to create new themes or theories, but to offer evaluations of how and why, akin to a case study.

Three central steps are needed to effectuate an audit trail using established standards and criteria to examine that the research project represents the phenomenon being explored: documentation for all methodological steps and analytical decisions, the auditor needs to be knowledgeable, an expert, and external with a clear contract, and provide appropriate records to conduct the audit (Cutcliffe, & McKenna, 1999; Halpern, 1983; Greene, 1988). Successful audit trails are planned from the beginning of the study, with data sources identified: planning and methodology, results, raw data, and analytical records (De Kleijn & Van Leeuwen, 2018). Audit trails, like other methods, should be systematic, named, and clearly reported processes and practices that often seem to be name-dropping with little erudition or disconfirmation (Coker, 2023).

Several problems exist with the audit trail. First, the researcher's inner thinking or a coherent, easily traced record might prove problematic if the qualitative research were truly recursive and iterative (Coker, 2026; Cutcliffe & McKenna, 1999). Secondly, De Kleijn and Van Leeuwen (2018) found that procuring truly independent experts who can challenge researchers can be difficult and must be negotiated prior to an audit. Thirdly, audit trails can scour the record for bias to evaluate overly smooth narratives or results, but audits can unintentionally contaminate results by introducing outside biases that defeat processes used in analysis (Bowen, 2009; Lietz & Zayas, 2010; Lillis, 1999). Defining clear objectives and reporting guidelines is nuanced and depends on the study's purpose and the qualitative methods employed.

Trustworthiness and credibility are key criteria for all qualitative research, yet researchers have found that many practices, such as audit trails, are rarely implemented and that there is limited substantive engagement with the research process (Carcary, 2009; Coker, 2023). Cutcliffe and McKenna (2004) called for a bold redevelopment of audit trails to intellectually challenge the status quo and create a vast improvement in qualitative research, such as an audit to test data to support propositions (Lillis, 1999) or an inquiry audit to examine, evaluate, and expand the scientific rigor (Bowen, 2009). Without a concerted, well-documented research process, sometimes lacking from qualitative research that spans years of starts and stops and does not retain or document all decisions, the downside is that an audit trail could detract from an otherwise rigorous study, as not all decisions can be reproducible, especially in interpretive research (Cutcliffe & McKenna, 2004).

Most 21st-century research offers little value or erudition on the aspects and implementation of audit trails. Many authors reiterate generalities rooted in the 1980s development of audit trails, offering little to no specificity beyond transforming Halpern's categories into broad areas of physical and intellectual activity (Bingham, 2023; Carcary, 2020; Earnest, 2020). Popular articles such as Ahmed (2024) and Richards and Hemphill (2018) provide no specific guidance on how to implement an audit trail or what needs to be done to effectuate one. Numerous examples exist beyond these researchers, but most offer little beyond Halpern's original research (Raskind et al., 2019; Lub, 2015). The gap is that most of the modern literature offers little specifics, parrots the original Halpern talking points, and lacks explication of actual practices in qualitative research across the disciplines (for example, see Lester et al., 2020, as but one of many examples). Popular textbooks provide only superficial appraisals and lack actionable methods and frameworks for implementing audit trails.

Researchers have long called for audit trails as necessary components of rigorous qualitative research to detail procedures to readers (Coleman, 2022). The problem is that it is unknown how researchers conduct audit trails in qualitative research. The research question, primarily descriptive, is as follows: How do researchers describe the audit trail method when reporting qualitative research in peer-reviewed articles? The purpose is to describe current audit trail methods and then use the results to improve the trustworthiness and credibility of qualitative research in general. A central concern is the rise of artificial intelligence (AI), which could be a tool for transforming current practices.

3. Methodology

Sampling criteria were that articles had to be strictly qualitative, participant-based research conducted through interviews; not meta-synthesis or content analysis; peer-reviewed; include a reported audit trail; be in English; have full text available; and be published from 2015 to 2025. The initial sampling size was projected to be approximately 50, which was considered sufficient for representativeness; however, the researcher would be open to increasing or reducing it depending on whether new directions were identified during data collection and analysis, which proceeded sequentially. Convenience sampling was used to identify general practitioners, though the sample may not be representative of all practices. Google Scholar was used to find and download peer reviewed articles by using search terms and combinations such as “audit trail,” “audit,” “qualitative,” “qualitative research,” etc.

The thematic analysis of content from peer reviewed articles was based on Boyatzis’s (1998) in an iterative and recursive manner. Step 1 began by downloading article references and relevant sections into Microsoft Word and concurrently familiarizing oneself with qualitative methodologies, methods, results, trustworthiness/credibility reporting, and the audit trail. In Step 2, the researcher checked line by line, given that the importance of each line was unknown. The coding schema was conducted in Microsoft Excel, with headings of author, title, website, and type of study, and then it proceeded with column headings of in vivo plus descriptive coding, and then the second cycle analysis considered where initial categories were formed by checking for patterns and clusters around frequencies, significant statements, and evaluative elements. Annotations and memos were used throughout, and the analysis was open to inductive coding when data did not fit predetermined categories or were deemed important. Step 3, the analysis began deductively to form categories, concepts, and descriptive summaries. The deductive categories examined audit trails according to Halpern's (1983) criteria. In Step 4, researchers developed descriptive themes to characterize the data, taking care to account for multivocality and to avoid a grand narrative that fails to account for outliers and disconfirmatory evidence. Descriptive themes were compared with the original data and evaluated to determine whether the results conformed to the participants' intent and purpose.

4. Results

Finding articles that employed audit trails was more time-consuming than envisioned. A search of over 3,000 articles yielded 39 articles. Many articles were in nursing, medicine, and healthcare; keywords related to education, leadership, and psychology were added to maximize variation. The results suggested that audit trails were more common in treatises on trustworthiness and credibility in qualitative research than in practice. The descriptions of audit trails in qualitative studies (Figure 1) were exceptionally brief: most write-ups were one or two sentences, with a range of one to two sentences the norm; on a few occasions, supplemental appendices reported a complete audit. Supplemental appendices were found twice in total, with complete renditions of the audits a stark exception.

Figure 1. Results of Analyzing Audit Trails in Qualitative Studies.

The central theme was that audit trails suffered from conceptual dilution: they lacked clear definitions, methodologies, or externalities, and their focus shifted from Halpern's dream of an examination and evaluation of methodological practices to offer recommendations for improvement to a stamp of approval. Confirmability and dependability, Halpern's idea, were mentioned, but the dimensionality and plausibility were always stated in favorable terms, as lack of transparency was ubiquitous. A minority of appendices followed Carcary rather than Halpern, offering a different take on audit trails. These auditors accepted advice and criticism unconditionally, though outsiders might question their willingness to bear witness. While some audit trails employed external evaluators, most researchers were either silent about who the auditors were or explicitly used team members to evaluate results.

What was the norm, besides the two exceptions, was that audit trails were guarantees that the results matched methodologies, methods, and results. No problems was the defining purpose. Technology was mentioned as enhancing audit trails, but the how and what were missing. Many audit trails were short and devoid of specifics; only a few researchers mentioned them in the abstract. There was no need for formal processes, and the lack of problems, questions, or explication of the how and what defined audit processes. Audit trails were frequently mentioned with other methods, such as prolonged engagement, reflexivity, and member checks, all of which also lacked formal methods and reporting of divergences.

Most audit trails served as proxies for quality without the erudition of methods. First, most audit trails were conducted internally to validate internal validity, as authors or the team of authors passed judgment on one's own work. Secondly, stating audit trails as a method without methods defined most reporting, with no mention of Halpern's ideas of preplanning, document/data generation and use, or specific purposes. Thirdly, technology usage served as a proxy, with assurances that technology created an auditable trail of evidence without erudition. The purpose was to guarantee results using the buzzwords "ensure," "assert," and "enhance," while purporting that the audit trail facilitated peers' acceptance and replication of results. The outliers, the two following Carcary, accepted audit trails as a means to move from evaluation and analysis to improving interpretations.

A possible limitation of this analysis is the word count requirements. Qualitative research often suffers from word counts that necessitate brevity, more apt for quantitative research than qualitative studies. Truly iterative and recursive qualitative research could involve elaborate, detailed methodologies, methods, and results that are neither linear nor a priori defined, but journals often do not permit such lengthy explanations. Notwithstanding this disclaimer, the drift from the original intent of audit trails in the 1980s to the present is quite distinct from what the purpose was envisioned.

5. Discussion

Coker (2023), reviewing frameworks for quality in qualitative research, found broadly similar results in qualitative dissertations for doctoral programs in the US: "Audits were popular in the dissertations, but most failed to mention any steps, and none used the original spirit promulgated by Halpern where there was a reevaluation, questioning, and negotiation of a study in its entirety" (p. 82). The future of audit trails can use the curious case of the typo in the title: Carcary (2009) accidentally titled the article about the audit trail as *audit trial*. The core problem with internal audits and reflexivity is that researchers must step outside their known world if problems, biases, and solutions are to be identified and addressed. When properly structured, generative AI can serve as a peer debriefer, confidant, and examiner, thereby providing all qualitative researchers with an external auditor. The proposal is that audit trails, as currently envisioned, drifted from the conceptual purpose and utility and should be reenvisioned with artificial intelligence (AI) as a tool to question, deconstruct, and improve practices. First, researchers must develop an AI framework for transparent usage. Secondly, researchers must determine the four types of audit trails that are reviewable, reportable, and potentially improvable. Recommendations and shortcomings are presented to assist all researchers in the ethical usage of AI in qualitative research.

AI use and debate have proliferated in the extant literature, with claims ranging from the dangers of AI use to claims that AI is as effective as, if not more effective than, humans. The debate here is that the current state of generative AI makes little sense, as generative AI will improve leaps and bounds for several years before plateauing; arguing over today would be of little benefit. A central proposition is that AI is widely used, from autocoding by most computer-assisted qualitative data analysis (CAQDAS) software to commercial generative AI programs such as ChatGPT, Google Gemini, Claude, and Perplexity, and a list seemingly growing by the day across all facets of qualitative research. The question is not if but when and how. The following steps are akin to peer review procedures, adapted from Halpern (1983) and De Kleijn and Van Leeuwen (2018), and provide a framework that could be adopted for the researcher-AI partnership:

Phase 1: Pre-Entry & Orientation

✓ Initiate contact by auditee
✓ Conduct orientation to the audit procedure
✓ Negotiate and agree upon goals, roles, and rules of the audit
✓ Discuss and review audit alternatives

Phase 2: Establishing Auditability

✓ Orientation to the study (review logistics and availability)
✓ Review audit trail components, structure, and linkages
✓ Determine auditability (assess for completeness, comprehensibility, and utility)

Phase 3: Formal Agreement & Contracting

✓ Negotiate the formal contract
✓ Establish timeline, goals, roles, and logistics
✓ Define product outcomes and format
✓ Establish criteria for renegotiation

Phase 4: Assessing Trustworthiness & Quality

Confirmability

✓ Assess if findings are grounded in data

✓ Verify that inferences are logical

✓ Evaluate the utility of the category structure

✓ Assess degree of inquirer bias (distinguish between disciplined vs. undisciplined bias)

Dependability

✓ Review appropriateness of inquiry decisions and methodological shifts

✓ Review bias used to bound the inquiry

✓ Assess overall design-implementation-integration of outcomes

Credibility & Visibility

✓ Review evidence of triangulation, debriefing, and member checks

✓ Assess the process for visibility, comprehensibility, and acceptability

✓ Assess corroboration between methodological choices, data sources, findings, and the audit trail

Phase 5: Renegotiation & Improvement (Peer Review Model)

✓ Auditor presents findings and discusses discrepancies

✓ Determine follow-up actions (redesigning research process, adjusting report, modifying agreement)

✓ Implement revisions and improvements

✓ Incorporate auditor recommendations

✓ Utilize integrative frameworks (e.g., COREQ) to guide activities

Phase 6: Closure

✓ Provide feedback

✓ Reach complete agreement

✓ Produce final written report and presentation

Researchers should implement the above phases with a transparent plan for using generative AI in qualitative research, as shown in Table 1. Stage one of the qualitative forensic audit trail tool-artificial intelligence (QFATT-AI) is the Transparent Reporting of AI Logistics (TRAIL) by following planned, reportable steps. Part of the process is that scientists should be recursive, iterative, and open to hiccups, unplanned excursions, and changes, all supposed properties of qualitative research often smoothed over in nice, neat, and clean narratives as if qualitative research were like solving a math problem with order of operations. TRAIL must be adopted and adapted for each stage and purpose of the QFATT-AI. The QFATT-AI is modular, so researchers can use one or more components as their research needs dictate. An effective QFATT-AI is developing an auditor contract to be implemented with a generative AI partner.

Table 1. Transparent Reporting of AI Logistics (TRAIL)

Steps	Definitions	Examples
Goal Specification: The Contract	**Goal**: Identifying the timing/nature of the audit (Pre-audit, Procedural Fidelity, In Situ/In Vivo, Post Hoc). **Purpose**: How the researcher uses the knowledge generated. **KPIs**: Key Performance Indicators to assess quality.	**Goal**: Post hoc. **Purpose**: To validate the consistency of manually derived themes. **KPI**: 90% inter-rater reliability agreement between AI and human coder."
Ethical Transparency	Confirm use of permitted with IRB or ethics approval.	"IRB approval included the use of generative models ….”
Evidence Verification	Substantiating computer-generated insights by quotes/raw data by original source.	"Themes identified validated with participant quotes (see Table 2)."
AI Plan	**Named:** Identifying tool. **Role**: * *Evaluation* (AI creates the process). *Analytical* (AI provides suggestions used by the researcher). *Interpretive* (AI as dialectical partner) **Task**: the omnibus objective **Parameters**: Specifications regarding length of results, writing style, and what material to exclude (not consider). **Evidence**: How data sources are utilized, findings are reported, and methods used detect/prevent AI hallucinations.	**Tool**: Claude 3.5. **Role**: Analytical—AI proposed initial codes which the researcher refined. **Task**: Summarize interview transcripts. **Parameters**: 200-word limit, academic tone, exclude non-verbal cues. **Evidence**: AI summaries were cross-referenced with audio files to ensure no hallucinations occurred."
Fair Reporting	Distinguishing between human and machine contributions, including timelines and percentage performed by each.	"The researcher manually coded 20% of the data to establish a baseline; the software processed the remaining 80%."
Data Governance (DS³)	**Data Source**: Identifying where the data originated. **Data Security**: Ensuring privacy, anonymity, and confidentiality. **Data Storage**: Confirming if AI retains data for training purposes.	"**DS³**: Data was sourced from anonymized transcripts (**Source**), processed on a local secure server (**Security**), and 'data retention' was disabled (**Storage**)."
Bias Mitigation	Acknowledging and addressing potential algorithmic biases inherent in the tool.	"Outputs were reviewed by two researchers to evaluate the tool did not misinterpret cultural idioms."

The TRAIL framework serves as the foundation for audit trails and for all AI use in qualitative research. The first step is the pre-audit, where researchers audit methodologies and plan the audit trail. Morse et al. (2002) described that a poor study cannot be claimed as trustworthy and credible regardless of claims:

> Of importance, an audit trail is of little use for identifying or justifying actual shortcomings that have impaired reliability and validity. Thus, they can neither be used to guide the research process nor to ensure an excellent product, but only to document the course of development of the completed analysis (Problem with, para 1).

A poor study is a poor study, regardless of the use or claims of an audit trail or other gimmickry, The first step of the QFATT-AI is the pre-audit to evaluate, analyze, and/or interpret all components of the methodology and methods before execution: theory/conceptual framework use, methodology, ethics plan, qualitative methods, population/sample frame/sample/sampling strategy, participant recruitment, data collection (e.g., raw data, interviews, focus groups, recording, transcriptions, etc.), data analysis (e.g., coding schema, memoing, annotating, etc.), trustworthiness/credibility criteria, and results reported (Bowen, 200). The omnibus goals are methodological coherence, non-biased and appropriate data collection from an optimal participant sample, and troubleshooting problems. Possible documents include Chapter 3 of a traditional five-chapter dissertation, the methodology section of a proposed research project, and notes, decisions, and discussions. Many existing research articles could assist, such as Coker and Akande's (2025) interview checklist, which is organized by problem and by global/granular levels, or the myriad frameworks reported by Coker (2023). If researchers plan to track decisions and biases, a formal method for constructing propositions throughout the research process is needed (Coker & Akande, 2025), as memory and anchoring biases can distort them, creating a false narrative.

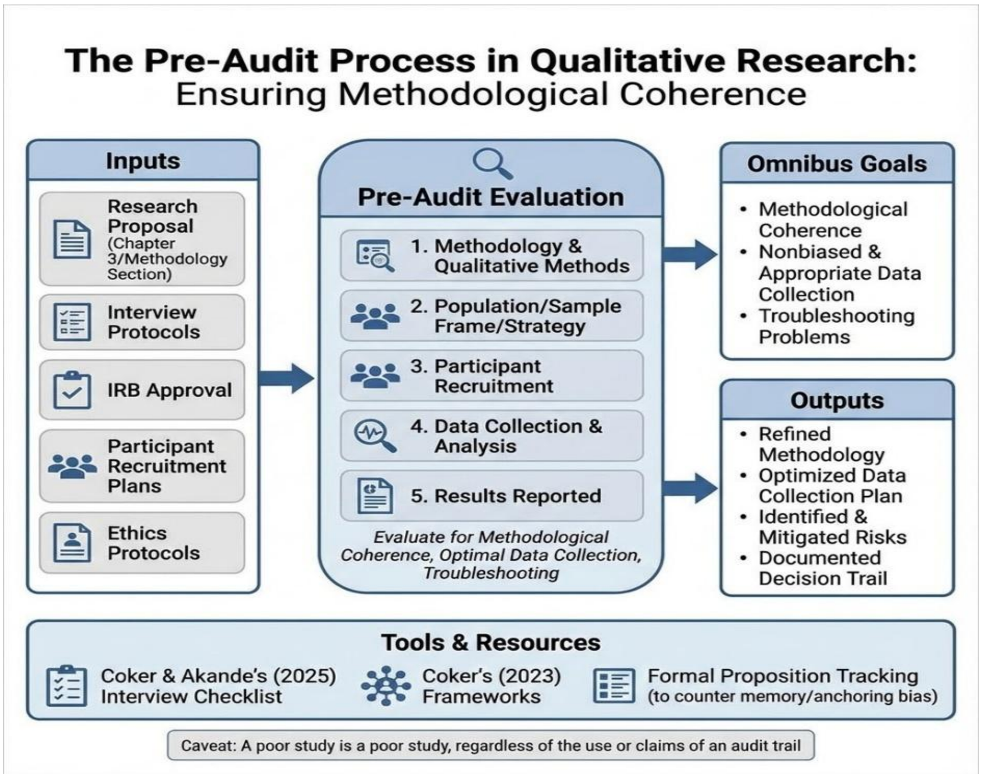

Figure 2. The Pre-Audit Process in the QFATT-AI

The procedural fidelity audit, shown in Figure 3, is akin to a checklist. In accordance with the TRAIL guidelines, the researcher requests an initial evaluation using the RPTE framework. The data source will be the planned methodology and the data from implemented plans, such as Chapters 3 and 4 of a dissertation, with the product akin to a checklist for comparing actual versus projected outcomes. The role is evaluation, to create a rubric of planned actions and goals of the methodology and methods, the role of a researcher evaluating if the methodology and methods match the actual practices, the task of comparing the before and after, parameters of an evaluative result of exceeds, meets, partially meets, problems, and/or omits, and evidence for each criteria. Researchers can enhance this stage by incorporating external resources, such as conducting thematic analysis using Braun and Clarke's checklists. Coker (2023) has consistently found that, due to the iterative, recursive nature of research, researchers often change plans or fail to implement methods without transparent reporting. The procedural audit can prompt the researcher to improve coherence and fidelity, particularly when data collection needs have ceased and are no longer required.

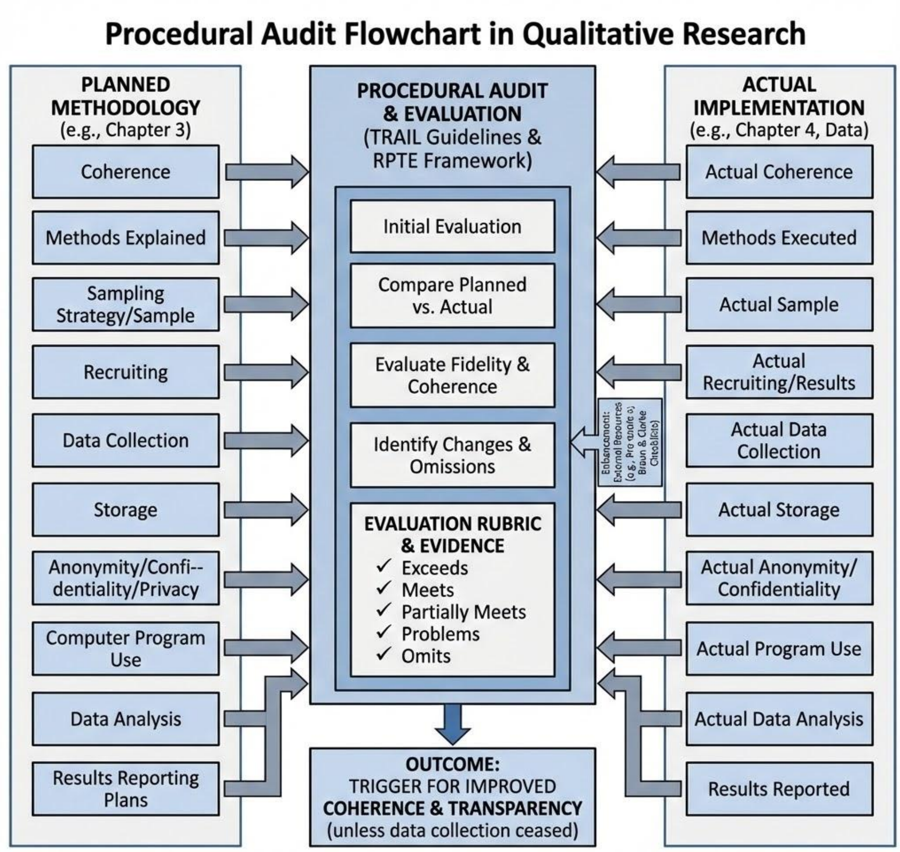

Figure 3. The Procedural Fidelity Audit in the QFATT-AI

The in situ/in vivo and post hoc audit trail utilize similar procedures but for different objectives. Both sets offer possibilities of evaluation, analysis, and/or interpretation of analysis and results to examine if the results are grounded in and supported by the data, though the actual analytic decisions are often missing (Halpern, 1983; Lincoln & Guba, 1985; Steccolini, 2023), but the in situ/in vivo is in the moment and causes a ripple effect of future research practices while post hoc is after the study and can lead to either a remodeling of results or reporting the evaluation for limitations. Halpern's (1983) becomes useful here, as researchers seek to move beyond procedural checklists and explore thoroughness and justification of data collection and analysis as well as the difference between researcher and pernicious bias. Data sources would be from data collection (raw data, transcripts, field notes, etc.), data reduction/analysis products (codes, memos, annotations, etc.), data reconstruction/synthesis products (rough drafts, final products, etc.), process notes (reflexivity journals, chronicles of thoughts, etc.), materials related to researcher (proposals, rough drafts, faculty meetings, researcher notes, etc.), instrument development (interview protocol rough drafts, pilots, subject matter expert use, etc.), verification/validity (processes for trustworthiness and credibility, etc.), and auditor processes (researcher writing decisions for the audit in anticipation and response, etc.) (Akkerman et al., 2008; Lincoln & Guba, 1985). Depending on whether the study is in situ/in vivo or post hoc, all steps of the research process are either carried out or completed.

The QFATT-AI, at this point, incorporates the Halpern criteria but in a way that honors Popper's idea of science as falsification of hypotheses and not the trick of verification or guarantees, as if the inherent post-positivist nature of subjectivity can be subverted into a positivist nature of extreme certainty and claims of ensure (Helfenbein & DeSalle, 2005). Coker and Liou (2025) found that positionality often seemed to determine results: no one could find critical race theory (CRT) proponents who ever found disproof of their preferred theory, and, just as ominously, opponents of CRT never found proof. The research questions, theory, and positionality could have been presented as results. The problem of generating preordained results is common: "Some common problems are fitting one's theory, positionality, personal motivation, perspective (going native), or research questions to the data rather than letting the data drive the results" (Coker, 2026).

Researchers should reject the problem that reflexivity, usually through journaling often not described or in generalities, can eliminate biases. Coker (2023) found that few researchers clearly articulated the meaning or value of reflexivity, casting doubt on the validity of their methods. Drawing strengths in numbers, many leading researchers drafted a letter stating 416 researchers—as if numbers for popularity, not logic or science, determine reliability and validity—rejected AI use in qualitative research: "We hold the position that only a human can undertake reflexive qualitative analytical work, and therefore use of GenAI is inappropriate in all phases of reflexive qualitative analysis, including initial coding" (Jowsey et al., 2025, p. 2). Nary will such a naïve, wrongheaded statement ever again be seen in the annals of qualitative research. As if a tool at hand in the Heideggerian experience were no more inappropriate than CAQDAS or video conferencing. At one time, all too were considered antithetical to conducting high-quality research. AI as a tool is here to stay, and the Jowsey et al. article is not only dated but also outdated. The idea that reflexivity could not be improved beyond the self through conversations and debates has not proven viable in a century of research, in which people are proficient at seeing the biases and blemishes of others but not their own (Pronin & Hazel, 2023).

If the reflexive racket does not work, what is to be done? Dror (2023) pointed out that experts are often more susceptible to error and bias than nonexperts, and that a systematic consideration of disruptive biases can improve forensic science more than turning inward or practicing without external constraints. The three Ds of reflexivity can be addressed by the QFATT-AI through a partner-in-researcher: doubts, dysfluency, and disconfirmation. Challenging oneself is paramount, but this is often difficult in human-to-human audit trails, where colleagues who either already agree or do not want to hurt the researcher's feelings are used. AI does not care about feelings, and neither should the researcher.

To develop in situ/in vivo or post hoc audit trails, three omnibus traits within the Dror framework can support the evaluation, analysis, and interpretation of qualitative research. First, researchers need to use paradoxical thinking to consider the opposite and the orthogonal, including simultaneous contradictory truths or possibilities (Calbretta et al., 2017). Secondly, the null hypothesis trick, in which the researcher has AI immerse itself in the data to disprove or reject the researcher's hypotheses or constructed themes, is one way to challenge possibilities during and after theme formation (Maher et al., 2018; Wesley, 2014). Thirdly, searching for negative cases directly challenges interpretations and offers more than reconstructing themes: dimensions, outliers, and multivocality can be explored (Chenail & Maione, 1997). The QFATT-AI omnibus framework, as shown in Figure 4, enables researchers to step outside their own perspectives and engage with those of people who disagree or view the world differently. Researchers can draw on a range of data sources, from raw data to other theories, findings, and questions. Technical work, such as sentiment analysis and keywords, can improve the audit, as well as the creativity of multiple themes, restorying, and artistic representations.

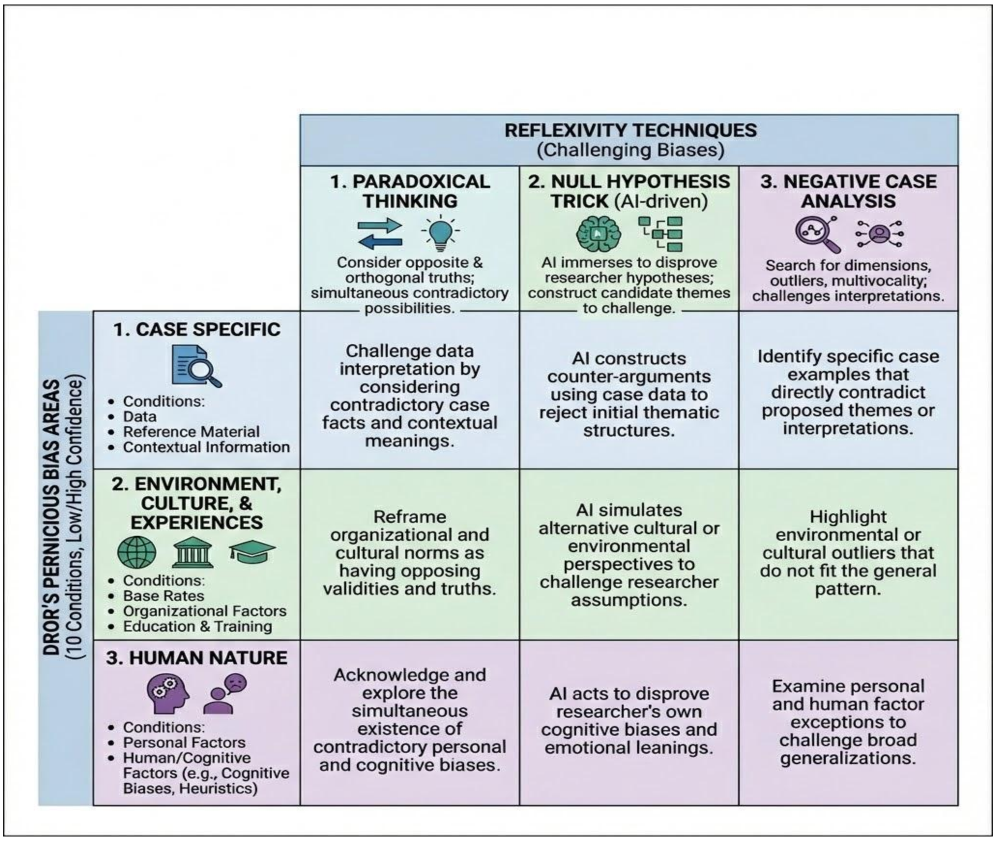

Figure 4. Reflexivity and Antibiasing Matrix

With the above set of conditions to challenge biases by paradoxical thinking, the null hypothesis trick, and negative case analysis, researchers must confront two types of biases: researcher instrumentality and pernicious bias. Researcher instrumentality means the researcher is an instrument, living in and recounting the world, and such experiences, vicarious and real, add richness and should be transparently reported beyond a laundry list of demographics in a positionality statement. The other concern is pernicious bias, in which researchers fail to address their cognitive biases regarding case-specific, individual-specific, and human-nature issues. Three areas are divided into 10 possible conditions that are often determined by whether the researcher has low or high confidence to define pernicious bias: 1.) Case-specific: data, reference material, contextual information; 2.) Environment, culture, and experiences: base rates, organizational factors, education, and training; 3.) Human nature: personal factors and human/cognitive factors (Dror et al., 2015; Dror et al., 2021).

A key is that if researchers are to create the three D’s, the AI must be programmed and directed to create an upside-down world. As shown in Table 2, using Halpern’s (1983) question-plus-reflexivity measures and Dror's, researchers can use the QFATT-AI within the TRAIL and RPTE formats to conduct evaluations, analyses, and interpretations. Avatars and personas are important components of the process. Avatars direct the AI tool to act in ways that embody Popper’s falsification mindset: Supporter, Detractor, Questioner, Expander, and Debiaser. Personas are another subset of avatars, such as asking AI to answer as if a 5th-grade student, supporting a position, or other hypotheticals. While traditional audit trails were used to confirm findings (Cutcliffe & McKenna, 2004), such an endeavor is impossible and should be replaced by plausibility, while informing readers of alternative possibilities.

Table 2. Operationalizing Avatars and Personas

Component	Type/Role	Operational Definition	Strategic Purpose (Popper’s Mindset)
Avatar	**Supporter**	Directs the AI to use the researcher’s specific inputs to validate and find a coherent direction.	**Verification:** Establishes the baseline argument or hypothesis to be tested.
Avatar	**Detractor**	Directs the AI to take a position totally against the researcher's findings or themes.	**Falsification**: Forces the researcher to confront the "upside-down world" and defend against total opposition.
Avatar	**Questioner**	Directs the AI to express doubt, skepticism, and demand more information or evidence.	**Doubts**: Identifies gaps in logic, missing data, or weak evidentiary links.

Avatar	**Expander**	Directs the AI to go on "what if" journeys, extending concepts beyond their current boundaries.	**Dimensions**: Explores outliers, alternative contexts, and orthogonal possibilities.
Avatar	**Debiaser**	Directs the AI to evaluate if biases are present (correct/incorrect) and identify a "third way."	**Disconfirmation**: Mitigates pernicious bias and researcher instrumentality.
Persona	**Identity Simulation**	Instructs the AI to adopt a specific viewpoint (e.g., Fellow Researcher, Participant, Reader, or a 5th Grade Student).	**Multivocality**: Shifts the analysis from confirmation of findings to "plausibility" by testing how different audiences interpret the data.

Five key areas, generated from Carcary (2020), Halpern (1983), Heath (1987), Miller (1987), and White et al. (2012), can guide in situ/in vivo or post hoc evaluation, analytics, and interpretations:

1. Foundational Design and Strategic Alignment

- Conceptual Clarity: Clearly define the research problem, questions, aims, and the gap in existing literature, ensuring the study design is congruent with these elements.
- Philosophical and Methodological Stance: Articulate the researcher's philosophical position and provide a transparent rationale for the chosen strategy (e.g., case study, design science) and sampling methods (e.g., purposive, snowball).
- Design Evolution: Document how the research design evolves as the process unfolds, including a record of all methodological decisions and shifts made over the course of the study.

2. Rigorous Data Collection and Management

- Literature Protocol: Establish and document a validated protocol for the literature review, including search scope, keyword strings, screening criteria, and quality appraisal standards.
- Evidence Sourcing: Explicitly specify all sources of primary data (e.g., transcripts, pilot feedback) and secondary evidence (e.g., policy documents, contextual descriptions).
- Instrument Development: Record the development and revision of research instruments, such as interview questions, and track the actual sequence of how data were collected and processed.

3. Analytical Transparency and Traceability

- Process Logic: Detail the procedures for data reduction (transcription, field notes) and reconstruction (category development, concept integration), ensuring the sequence from data to conclusion is traceable.
- Thematic Mapping: Document all codes and categories, using memos to trace emerging themes back to the body of evidence and diagrammatically exploring relationships across the thematic structure.
- Evidence Linking: Ensure conclusions are explicitly linked with exhibits of condensed data, demonstrating that findings are grounded in the evidence rather than researcher assumption.

4. Reflexivity and Researcher Positionality

- Bias and Assumption Disclosure: Maintain a record of the researcher's personal assumptions, values, affective states, and "hunches" (preliminary hypotheses) to monitor their influence on the study.
- Role and Experience: Explicitly describe the researcher's role and status within the site, as well as how prior personal and professional experiences have influenced the conceptualization of the research.
- Intellectual Transparency: Keep a reflexive journal to document motivations, reactions to informants, and the transparency of analytical thinking during the interpretation phase.

5. Verification and Quality Assurance

- Validation Strategies: Implement and document systematic procedures for reliability and validity, such as triangulation, member checking, and peer debriefing.
- Auditability Assessment: Maintain records detailed enough to allow an external auditor to determine the completeness, comprehensibility, and utility of the process.
- Dissemination Strategy: Specify the target audience and ensure the research report is documented in a manner transparent to that audience, discussing findings vis-à-vis prior studies

The modular format of audit trails is an important component, as a limited audit trail that is neither iterative nor recursive can corrupt the process of meeting external guidelines regardless of optimal practices (Greene et al., 1988). Reflexivity can be intersubjective, where the researcher interacts with participants and data to create a third way (LaBanca, 2011), as biases can affect the entire study, from research questions and interview questions to how results are framed (Johnson et al., 2020). Krefting (1991), citing Lincoln and Guba, discussed the reflexive log as a means to document the researcher's journey and as a potential record of biases, but assumes that researchers can actively seek out and identify their own biases. The idea that researchers can succeed where others have failed lacks strong support in psychology or the cognitive sciences. The Halpern (1983) dissertation, supported by Greene et al. (1988), as presented in Appendix A, provides additional material for generating the QFATT-AI.

Audit trails should be conducted in phases and address the what and why of the researcher before, during, and after a research study (Mertens et al., 2025). Calls for self-reflection have long been used in qualitative research (Creswell, 2003), but the how was often left to chance. Bias in research is inherent and does not necessarily detract from practices and results (Cutcliffe & McKenna, 2004), but separating researcher bias from pernicious bias is more than a researcher journaling in the hopes of an aha moment. The major advantage of the QFATT-AI is saving time spent creating and charting summaries, and providing an evaluator, cooperator, collaborator, and co-researcher who will challenge and debate the researcher.

6. Conclusion

Reflexive qualitative research practices promote audit trails to develop transparency (Nicmanis, 2024), but without moving beyond the favored inference with concrete reasons, the audit trail may be of little value (Wesley, 2014). The downside in such reflexive practices is the problem of performativity: If a researcher knows his/her notes will reflex on the self, it is a difficult scenario to believe the researcher would not consider social desirability even if not overtly mentioned. No one wants to appear like a rube. Also, if someone knew they were biased or wrong, the person would not believe what they purport to believe.

Positionality statements and traditional reflexivity have gained widespread popularity, but the dam is beginning to break due to their lack of efficacy (Savolainen et al., 2023). Cutcliffe and McKenna (2004), citing Kvale, point out a key problem with audit trails: seeking legitimation through transparency rather than judging the findings by their value. While Akkerman et al. (2008) believe audit trails keep people honest, the problem lies in the aforementioned performativity, the substitution of methods for results, and the lack of audits that allow reporting beyond a mere stamp of success. The QFATT-AI will not correct poor data instrumentation or unethical behavior—neither has existing structure—but can provide the lone researcher or a research team a chance to step outside their own skin and assess their processes and results in a systematic, rigorous manner. Dado et al. (2023) claim that reflexive journaling and audits are necessary, but current practices suggest these are empty gestures, devoid of systematic methods and almost wholly lacking in reporting problems and shortcomings. The QFATT-AI offers researchers a chance to step outside their own perspectives, consider new angles, and develop a dialectic to challenge assumptions and beliefs.

A Call to Action

1. Adopt the QFATT-AI Framework. Move beyond static, performative checklists by integrating the Qualitative Forensic Audit Trail Tool—Artificial Intelligence (QFATT-AI). Use this modular system not just to verify findings, but to actively interrogate your data, methodology, and results through a rigorous, forensic lens.

2. Implement the TRAIL Protocol. Commit to radical transparency by using the Transparent Reporting of AI Logistics (TRAIL). Explicitly document every instance of AI usage, from the specific tool and prompt parameters to the percentage of human-vs-AI contribution, to ensure your study adheres to ethical standards and remains auditable.

3. Engage in Upside-Down Thinking. Shift your analytical goal from confirmation to falsification. actively employ AI avatars (such as the Detractor or Questioner) to deploy the null hypothesis trick and negative case analysis, forcing you to defend your themes against generated counter-arguments and orthogonal possibilities.

4. Operationalize Reflexivity. Transform reflexivity from a passive journaling exercise into a dynamic dialectical process. Use AI as an external critical friend to expose pernicious bias (case-specific or cognitive errors) that you cannot see yourself, rather than relying on colleagues who may hesitate to critique your work.

5. Prioritize Plausibility Over Guarantees. Stop using audit trails as a stamp of approval or a guarantee of trustworthiness. Instead, use the audit process to demonstrate plausibility by transparently reporting the messy, iterative reality of your research, including the divergences, problems, and alternative explanations that were considered and rejected.

References

Ahmed, S. K. (2024). The pillars of trustworthiness in qualitative research. *Journal of Medicine, Surgery, and Public Health*, *2*, 100051. https://doi.org/10.1016/j.glmedi.2024.100051

Akkerman, S., Admiraal, W., Brekelmans, M., & Oost, H. (2008). Auditing quality of research in social sciences. *Quality & Quantity, 42*(2), 257–274. http://dx.doi.org/10.1007/s11135-006-9044-4

Bingham, A. J. (2023). From data management to actionable findings: A five-phase process of qualitative data analysis. *International Journal of Qualitative Methods*, *22*, 16094069231183620. https://doi.org/10.1177/16094069231183620

Bowen, G. A. (2009). Supporting a grounded theory with an audit trail: An illustration. *International Journal of Social Research Methodology, 12*(4), 305–316. https://doi.org/10.1080/13645570802156196

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development.* Sage.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101. https://www.tandfonline.com/doi/abs/10.1191/1478088706QP063OA

Calabretta, G., Gemser, G., & Wijnberg, N. M. (2017). The interplay between intuition and rationality in strategic decision making: A paradox perspective. *Organization Studies, 38*(3-4), 365–401. https://doi.org/10.1177/0170840616655483

Carcary, M. (2009). The research audit trial—enhancing trustworthiness in qualitative inquiry. *Electronic Journal of Business Research Methods, 7*(1), 11–24. https://academic-publishing.org/index.php/ejbrm/article/view/1239/1202

Carcary, M. (2020). The research audit trail: Methodological guidance for application in practice. *Electronic Journal of Business Research Methods*, *18*(2), 166-177. https://doi.org/10.34190/JBRM.18.2.008

Chenail, R. J., & Maione, P. (1997). Sensemaking in clinical qualitative research. *The Qualitative Report, 3*(1), 1–10. https://doi.org/10.46743/2160-3715/1997.2030

Coker, D. C. (2021). Making thematic analysis systematic: The seven deadly sins. *Journal of Studies in Education, 11*(3), 126–146. https://doi.org/10.5296/jse.v11i3.18882

Coker, D. C. (2023). An integrative qualitative framework: Improving research through strategic mapping. *International Research in Education*, *11*(1), 66-109. https://doi.org/10.5296/ire.v11i1.20921

Coker, D. C. (2026). *Building the thematic structure in qualitative research: The theme construction tool.* [Manuscript submitted for publication]. College of Education and Human Performance, West Liberty University.

Coker, D. C., & Akande, A. (2025). Preparation, development, and refinement of the interview protocol in qualitative research. *The Qualitative Report*, *30*(8), 4205-4235. https://doi.org/10.46743/2160-3715/

Coker, D. C., & Liou, Y. H. (2025). Leadership research redefined: Exploring effectiveness, cross-cultural engagement, and social networking. In A. Akande (Ed.), *Power, politics and influence: Exercising followership, leadership, and practicing politics* (pp. 103-131). Cham: Springer Nature Switzerland.

Coleman, P. (2022). Validity and reliability within qualitative research for the caring sciences. *International Journal of Caring Sciences*, *14*(3), 2041-2045. http://www.internationaljournalofcaringsciences.org/

Cutcliffe, J. R., & McKenna, H. P. (1999). Establishing the credibility of qualitative research findings: The plot thickens. *Journal of Advanced Nursing, 30*(2), 374–380. https://doi.org/10.1046/j.1365-2648.1999.01090.x

Cutcliffe, J. R., & McKenna, H. P. (2004). Expert qualitative researchers and the use of audit trails. *Journal of Advanced Nursing, 45*(2), 126–133. https://doi.org/10.1046/j.1365-2648.2003.02874.x

Dado, M., Spence, J. R., & Elliot, J. (2023). The case of contradictions: How prolonged engagement, reflexive journaling, and observations can contradict qualitative methods.

International Journal of Qualitative Methods, *22*, 16094069231189372. https://doi.org/10.1177/16094069231189372

De Kleijn, R., & Van Leeuwen, A. (2018). Reflections and review on the audit procedure: Guidelines for more transparency. *International Journal of Qualitative Methods, 17*(1), 1609406918763214. https://doi.org/10.1177/1609406918763214

Dror, I. E. (2023). The most consistent finding in forensic science is inconsistency. *Journal of Forensic Sciences*, *68*(6), 1851-1855. https://doi.org/10.1111/1556-4029.15369

Dror, I., Melinek, J., Arden, J. L., Kukucka, J., Hawkins, S., Carter, J., & Atherton, D. S. (2021). Cognitive bias in forensic pathology decisions. *Journal of Forensic Sciences*, *66*(5), 1751-1757. https://onlinelibrary.wiley.com/doi/pdfdirect/10.1111/1556-4029.14697

Dror, I. E., Thompson, W. C., Meissner, C. A., Kornfield, I., Krane, D., Saks, M., & Risinger, M. (2015). Context management toolbox: A linear sequential unmasking (LSU) approach for minimizing cognitive bias in forensic decision making. *Journal of Forensic Sciences*, *60*(4), 1111-1112. https://doi.org/10.1111/1556-4029.12805

Earnest, D. (2020). Quality in qualitative research: An overview. *Indian Journal of Continuing Nursing Education*, *21*(1), 76-80. https://doi.org/10.4103/IJCN.IJCN_48_20

Fetterman, D. M. (1988). Qualitative approaches to evaluating education. *Educational Researcher, 17*(8), 17–23. https://doi.org/10.3102/0013189X017008017

Gerson, E. M. (1984). Qualitative research and the computer. *Qualitative Sociology, 7*(1), 61–74. https://doi.org/10.1007/BF00987107

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report*, *8*(4), 597-607. https://doi.org/10.46743/2160-3715/2003.1870

Greene, J. C., Doughty, J., Marquart, J. M., Ray, M. L., & Roberts, L. (1988). Qualitative evaluation audits in practice. *Evaluation Review, 12*(4), 352–375. https://doi.org/10.1177/0193841X8801200402

Halpern, E. S. (1983). *Auditing naturalistic inquiries: The development and application of a model* [Unpublished doctoral dissertation]. Indiana University.

Heath, A. W. (1997). The proposal in qualitative research. *The Qualitative Report, 3*(1), 1–4. https://doi.org/10.46743/2160-3715/1997.2026

Helfenbein, K. G., & DeSalle, R. (2005). Falsifications and corroborations: Karl Popper's influence on systematics. *Molecular Phylogenetics and Evolution*, *35*(1), 271-280. https://doi.org/10.1016/j.ympev.2005.01.003

Johnson, J. L., Adkins, D., Chauvin, S. W., & Carey, W. C. (2020). Qualitative research in pharmacy education: A review of the quality indicators of rigor in qualitative research. *American Journal of Pharmaceutical Education*, *84*(7), Article 7798. https://doi.org/10.5688/ajpe7798

Jowsey, T., Braun, V., Lupton, D., & Fine, M. (2025). We reject the use of generative artificial intelligence for reflexive qualitative research. *SSRN*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5676462

Korstjens, I., & Moser, A. (2017). Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing. *The European Journal of General Practice*, *24*(1), 120–124. https://doi.org/10.1080/13814788.2017.1375092

Krefting, L. (1991). Rigor in qualitative research: The assessment of trustworthiness. *The American Journal of Occupational Therapy, 45*(3), 214–222. https://doi.org/10.5014/ajot.45.3.214

LaBanca, F. (2011). Online dynamic asynchronous audit strategy for reflexivity in the qualitative paradigm. *The Qualitative Report, 16*(4), 1160–1171. https://doi.org/10.46743/2160-3715/2011.1125

Leko, M. M., Cook, B. G., & Cook, L. (2021). Qualitative methods in special education research. *Learning Disabilities Research & Practice*, *36*(4), 278-286. https://doi.org/10.1111/ldrp.12268

Lester, J. N., Cho, Y., & Lochmiller, C. R. (2020). Learning to do qualitative data analysis: A starting point. *Human Resource Development Review*, *19*(1), 94-106. https://doi.org/10.1177/1534484320903890

Lietz, C. A., & Zayas, L. E. (2010). Evaluating qualitative research for social work practitioners. *Advances in Social Work, 11*(2), 188–202. https://doi.org/10.18060/589

Lillis, A. M. (1999). A framework for the analysis of interview data from multiple field research sites. *Accounting & Finance, 39*(1), 79–105. https://doi.org/10.1111/1467-629X.00018

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE.

Lub, V. (2015). Validity in qualitative evaluation: Linking purposes, paradigms, and perspectives. *International Journal of Qualitative Methods*, *14*(5), 1609406915621406. https://doi.org/10.1177/1609406915621406

Maher, C., Hadfield, M., Hutchings, M., & de Eyto, A. (2018). Ensuring rigor in qualitative data analysis. *International Journal of Qualitative Methods*, *17*(1). https://doi.org/10.1177/1609406918786362

Megheirkouni, M., & Moir, J. (2023). Simple but effective criteria: Rethinking excellent qualitative research. *The Qualitative Report*, *28*(3), 848-864. https://doi.org/10.46743/2160-3715/2023.5845

Mertens, B., De Maeyer, S., & Donche, V. (2025). Structured audit trail for making the rigor in qualitative research explicit: A methodological tutorial. *SocArXiv* (prepint). https://doi.org/10.31235/osf.io/9qekw_v1

Miller, D. L. (1997, March 24-28). *One strategy for assessing the trustworthiness of* *qualitative research: Operationalizing the external audit* [Paper presentation]. Annual Meeting of the American Educational Research Association, Chicago, IL, United States. https://files.eric.ed.gov/fulltext/ED411268.pdf

Morse, J. M., Barrett, M., Mayan, M., Olson, K., & Spiers, J. (2002). Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods*, *1*(2), 13-22. https://journals.sagepub.com/doi/full/10.1177/160940690200100202

Nicmanis, M. (2024). Reflexive content analysis: An approach to qualitative data analysis, reduction, and description. *International Journal of Qualitative Methods*, *23*, 16094069241236603. https://doi.org/10.1177/16094069241236603

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods, 16*(1). https://doi.org/10.1177/1609406917733847

Owens, R. G. (1982). Methodological rigor in naturalistic inquiry: Some issues and answers. *Educational Administration Quarterly, 18*(2), 1–21. https://doi.org/10.1177/0013161X82018002002

Pronin, E., & Hazel, L. (2023). Humans' bias blind spot and its societal significance. *Current Directions in Psychological Science*, *32*(5), 402-409. https://doi.org/10.1177/09637214231178745

Ranaweera, N. (2024). Beyond numbers: Embracing the depth of qualitative research in criminology. *International Journal of Qualitative Research*, *3*(3), 315-325. https://doi.org/10.47540/ijqr.v3i3.1346

Raskind, I. G., Shelton, R. C., Comeau, D. L., Cooper, H. L., Griffith, D. M., & Kegler, M. C. (2019). A review of qualitative data analysis practices in health education and health behavior research. *Health Education & Behavior*, *46*(1), 32-39. https://doi.org/10.1177/1090198118795019

Richards, K. A. R., & Hemphill, M. A. (2018). A practical guide to collaborative qualitative data analysis. *Journal of Teaching in Physical Education*, *37*(2), 225-231. https://doi.org/10.1123/jtpe.2017-0084

Savolainen, J., Casey, P. J., McBrayer, J. P., & Schwerdtle, P. N. (2023). Positionality and its problems: Questioning the value of reflexivity statements in research. *Perspectives on Psychological Science*, *18*(6), 1331-1338. https://doi.org/10.1177/17456916221144988

Steccolini, I. (2023). What counts as “good” qualitative accounting research? Researchers' perspectives on assessing and proving research quality. *Accounting, Auditing & Accountability Journal*, *36*(3), 1032-1057. https://doi.org/10.1108/AAAJ-05-2022-5808

Weil, S. (2008). Is there a "legitimation crisis" in qualitative methods? *Forum: Qualitative Social Research*, *9*(2). https://doi.org/10.17169/fqs-9.2.438

Wesley, J. J. (2014). The qualitative analysis of political documents. In E. G. Miller & B. C. Müller (Eds.), *From text to political positions: Text analysis across disciplines* (pp. 135–160). John Benjamins Publishing Company. https://doi.org/10.1075/dapsac.55.07wes

White, D. E., Oelke, N. D., & Friesen, S. (2012). Management of a large qualitative data set: Establishing trustworthiness of the data. *International Journal of Qualitative Methods, 11*(3), 244–258. https://doi.org/10.1177/160940691201100305

Appendix A

Halpern's (1983) Guidelines for the Audit Trail

Purpose	Objective	Tasks, Actions, & Modern Integrations
Orientation & Contract	Establish Auditability	Negotiate: Define timeline, logistics, and nature of the study. Alignment: Ensure auditor's background matches Research Questions (RQs). Sampling: Define audit sampling strategy (auditor cannot reconstruct all). Suitability: Assess availability of evidence (raw data, notes) for audit.
Dependability	Process Fidelity	Plan vs. Action: Compare proposed methodology against actual execution. Rationale: Evaluate support for decisions regarding material/protocol changes. Adherence: Verify strict adherence to interview and sampling protocols.
Bias Detection	Inquirer Integrity	Early Closure: Identify unexplored areas in raw data (Negative Case Analysis). Cooptation: Detect overuse of personal notes over participant data. Premature Judgment: Spot unfounded convergence of opinion and field data. Challenge: Question "Inquirer Naivete" (unnecessary adherence) in debriefs. AI Integration: Deploy AI agents to simulate "Detractor" personas.
Confirmability (Linkage)	Verify Isomorphism	Chain of Evidence: Trace logic: Raw Data to Reduction to Findings. Grounding: Ensure all conclusions anchor to specific data points. Logic Check: Flag illogical inferences or unsubstantiated claims. Reasoning: Document use of abductive reasoning vs. linear deduction. Omissions: Flag unused data or unexplained conflicting evidence.
Confirmability (Structure)	Categorical Structure	Optimal Set: Ensure categories are mutually exclusive and exhaustive. Clarity: Assess thematic overlap and level of abstraction. Alternatives: Document consideration/rejection of alternative categories. Terminology: Ensure researcher did not impose own terms on data. Testing: Verify iterative theme testing (Test-Operate-Test).
Auditability (Replay)	Mental Process	Process Notes: Examine methodological notes, decisions, and rationale. Evolution: Trace the chronological evolution of the study design. Replay: Ensure notes allow auditor to cognitively "replay" the analysis. Intentions: Review personal notes to separate motivation from findings.
Closure & Synthesis	Iterative Reporting	Debriefing: Conduct sessions where questions lead to immediate refinement. Correction: Allow researcher to correct deficiencies before final report. Synthesis: Report degree of adherence to trustworthiness guidelines. AI Trial: Run a "Hostile Audit" simulation on final structure before publishing.

Acknowledgments

Not Applicable.

Funding

Not Applicable.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Macrothink Institute.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.