

# Causal mapping is an interesting QDA approach which is very suitable for scaling with AI

📅 25 May 2026

## Abstract

This paper argues that AI can be used inside a qualitative workflow without becoming the analyst, provided it is based on a method whose coding task is narrow enough to be locally checkable. Causal mapping is such a method: the basic coded unit is a single causal claim made by a source, grounded in a verbatim quote. We describe a workflow in which the AI does one narrow extraction task, pairing each candidate causal link with the exact quote that supports it, leaving almost no room for a black box, while every other analytic decision remains with the human researcher. The vocabulary of factors, the structure of the codebook, the analytic pipeline and the narrative interpretation do not need to be delegated to the AI. The workflow is exploratory in spirit. The AI identifies candidate causal claims, each tied to a quote, and the interpretive work of reading, comparing, reorganising and challenging those candidates still belongs to the analyst. We argue that this division of labour is a clean answer to the call for AI-assisted qualitative work that remains rigorous, transparent and accountable, and that the small but established tradition of causal mapping is well placed to support it.

See also:

- [Causal mapping as causal QDA](#) (companion paper, more detail on the QDA framing)
- [Minimalist coding for causal mapping](#) (the coding stance)
- [Combining opposites, sentiment](#) (label-level extensions)

## 1. Introduction

The arrival of large language models in qualitative research is producing two reasonable responses. The first is a worry that AI will erode the interpretive practices that give qualitative work its value: close reading, reflexive judgement, a rejection of positivism. The second is enthusiasm for the prospect of working at scales that were previously impractical. Both responses are warranted. The interesting question is whether a workflow can be designed that takes the productivity gains seriously while preserving the methodological commitments.

We argue that one such workflow already exists in causal mapping, a well-established family of approaches in evaluation and applied social research for representing what influences what according to sources (Axelrod 1976; Eden 1988; Powell et al. 2024; Evaluation 2024). The argument has two steps. The first introduces causal mapping as a form of QDA in its own right, sometimes called causal QDA, with a single distinctive coding act and a structured intermediate product. The second shows why this coding act, narrow and locally checkable as it is, is well suited to bounded AI assistance without surrendering judgement to a black box.

The argument is addressed to a methodologically mixed audience: people interested in language, discourse and interaction, but also in the practical question of how to use AI in qualitative work without losing the qualitative.

We should be clear about scope. This is closer to a small-Q than a big-Q argument. We do not engage the wider question of what counts as the AI's positionality, or whether a model can be a participant in interpretation in the same sense that an analyst is. We are not proposing causal QDA as a heroic method for identifying new paradigms in qualitative inquiry. The contribution is more modest: a way of getting useful work done for researchers and evaluators whose questions are about what other people think causes what, where the alternative is either much slower manual coding or a black-box synthesis that is hard to audit.

The workflow does not replace the reflexive, interpretive work that gives big-Q qualitative research its distinctive contribution. It runs alongside that work, and it is most useful when the analyst still wants to read closely but also wants help finding the passages worth reading closely. The aim is to make a corpus explorable rather than to package it as a finding.

## 2. Causal mapping as practically focused QDA

In ordinary qualitative coding a code typically denotes a theme or concept. In causal mapping the basic coded unit denotes a causal claim made by a source: a cause label, an effect label, a verbatim quote that supports the claim, and a source identifier. A coding act yields an ordered pair, **Cause** -> **Effect**, attached to evidence and provenance. The dataset of such acts is a links table, and that links table is the core qualitative product rather than a stepping stone to a narrative.

This is a small unit, but by design. We do not code strength, polarity, necessity, sufficiency, role as moderator or any of the other features that systems-dynamics and grounded-theory traditions sometimes attach to links (Kim & Andersen 2012). Most respondents do not state those features explicitly, and most analysts cannot reliably extract them from text. By holding the coded unit to bare causation, we preserve the chain of evidence and keep the act of coding within reach of both human coders and AI assistance. We have called this stance minimalist or barefoot coding (see [Minimalist coding for causal mapping](#)).

The links table can be queried directly. Every node is a factor that appears as a cause or effect in at least one claim. Every edge is a claim with a quote behind it. Natural questions include: which factors are the most frequently mentioned upstream influences on a given outcome; how pathways into a target factor differ across subgroups; which links are contested, with both  $X \rightarrow Y$  and  $X \rightarrow \sim Y$  appearing in different sources. Each question is answered by an explicit, reversible operation on the table, and each answer remains traceable to the underlying quotes. The full taxonomy of operations and worked analysis pipelines is developed in a companion methods paper (see [Causal mapping as causal QDA](#)); here we keep just enough of the methods to make the AI argument.

This makes causal QDA practically focused in a specific sense. The research question is causal, the coded unit is a causal claim, and the analysis is an explicit pipeline of operations on a structured intermediate product. Compared to a thematic analysis (Braun & Clarke 2021), it gives up some breadth of interpretive scope in exchange for a much tighter audit trail from any map back to the original quotes. For research questions about drivers, barriers, mechanisms and pathways, the trade is usually a good one; for worked examples in evaluation practice see (Remnant et al. 2025; Powell et al. 2025).

## 2.1 A worked example

If an interviewee says,

After the clinic started opening on Saturdays I did not have to miss work, so I could actually attend.

a thematic pass might code **Access**, **Clinic opening hours**, **Employment constraints** and **Attendance**. A causal-coding pass records two links:

- **Saturday opening** -> **Not missing work**
- **Not missing work** -> **Attendance**

Both representations have value. Only the causal one is queryable as a mechanism. A reader can ask which other factors point into **Attendance**, which contexts mention **Saturday opening**, which paths run from clinic operations to attendance, and so on, and every answer remains line-by-line accountable to the original quotes.

## 3. Where AI normally goes wrong in qualitative work

The standard worry about AI in qualitative analysis is about synthesis tasks. "Find the main themes in this corpus" or "Summarise what these interviewees say about X" are open-ended prompts whose output is sensitive to the model's implicit theory of what counts as a theme. The output may read well and may even be roughly right, but it is hard to audit. The model may have downweighted minority views, smoothed over disagreements, drifted off the text, or invented categories that fit its training distribution better than the data. Worse, when the analyst inspects the output, the apparent fluency tends to disarm scrutiny. Or — it might not have wandered off. But how would we know?

There is a second failure mode that is less often noticed. Even when the AI is asked to do something more constrained, such as "code each excerpt against this codebook", the inputs are usually long enough that the model is making implicit selection judgements about what to attend to. Two passes over the same text can produce different codings, not because the model is faulty, but because the task as posed leaves enough latitude for it to vary.

These observations are not arguments against using AI in qualitative work. They are arguments against giving the AI the analytic burden. Once we accept that AI does best on narrow tasks with locally checkable outputs, the question becomes: which qualitative methods have a coding act with that shape?

## 4. The causal coding task is the right shape for AI

The minimalist coding act in causal mapping is exactly that kind of narrow, locally checkable task. The instruction to the model is:

Identify each passage where the text says that one thing influenced another. For each, record the cause, the effect and the exact quote that supports the claim.

This instruction has several useful properties. It refers to features of the text that are already present, namely explicit causal claims. It produces outputs whose unit, a link with a quote, is easy to verify by reading the quote. It does not ask the model to weigh, summarise or theorise. Because each link is a separate unit, errors are local: a wrong link can be removed or corrected without unravelling the rest of the analysis.

In practice we run this extraction on short chunks of text, often a single passage at a time, and we recover the corpus-level structure by aggregating the resulting links. The model is not asked to hold the whole corpus in attention or to decide what matters across documents. It does the same small extraction job, repeatedly, on chunk after chunk.

When the output is wrong we usually do not correct individual links. We go back to the prompt, adjust it, recode and iterate. Because the unit cost of a coding pass is low and the output is locally checkable, this experimentation is cheap, and the analyst stays in control of what the AI is being asked to do.

The extracted links are candidates. They are raw material for analysis, ready to be read closely, compared, followed up or set aside. The analyst's job is to do that work: to follow surprising claims, drop dull ones, reorganise the vocabulary where useful, and use the resulting map as an instrument for exploring the corpus rather than as a finding to report.

This way of working is a qualitative variant of the split-apply-combine strategy (Wickham 2011). The split is the minimalist coding act; the apply step is a deterministic pipeline of operations on the links table; the combine step is human-authored synthesis. The clean mapping between strategy and method is what makes the AI question tractable.

## 5. AI as clerk, human as architect

A useful division of labour follows. The AI is a clerk: fast, consistent, willing to apply a stable rule across thousands of passages, and locally accurate enough for the rule to be worth applying (n.d.). The human is the architect, responsible for everything that requires judgement.

The architect's work includes, at minimum, the following.

- **Framing the research question.** What kind of causal claims matter for this study, and at what level of detail.
- **Choosing the factor vocabulary.** We curate the labels rather than letting the model invent freely. Where the model proposes labels, we cluster and rewrite using a separate operation called magnetisation (see [Magnetisation](#)), which is itself a deterministic transform on the links table.
- **Designing the hierarchy.** Where some factors are best read as instances of more general ones, we encode this in the labels themselves using a ; separator. This is a coding-time decision, not an AI inference.
- **Marking opposites.** Where the data contains paired factors such as employment and unemployment, we mark them as opposites using a tilde convention and let an explicit transform combine evidence across poles when useful. This too is a human-controlled label-level decision (see [Combining opposites, sentiment](#)).

- **Building the analytic pipeline.** Which filters to apply, which evidence thresholds to set, which paths to trace, how to render the resulting view. Every step is explicit and reproducible.
- **Writing the interpretation.** The narrative account of what the evidence shows, including its limits, the contested claims and the cases that do not fit.

None of this is delegated to the model in our workflow. The consequence is that the AI never produces an analytic claim. It produces candidate evidence in a fixed format, which the human then assembles into claims.

### 5.1 What is and is not in the AI's hands

It is useful to be very explicit about the boundary, because this is where critiques of AI-assisted qualitative work tend to land. In our workflow the AI sees one chunk of text at a time and is asked to extract causal claims from that chunk. It does not:

- choose which research questions matter,
- choose which sources to include,
- decide what counts as a factor in this study,
- decide where opposites or hierarchies are warranted,
- decide which paths through the map are interesting,
- summarise across documents,
- write any part of the final account.

It does:

- propose candidate cause and effect labels for each apparent causal claim in the chunk,
- pair each candidate link with the exact quote that supports it,
- optionally flag uncertain cases for human review.

Every analytic move in the rest of the workflow is either the analyst's direct authorship or a deterministic transform whose code can be inspected. The AI is one bounded component of a much larger, mostly non-AI process.

## 6. What auditability looks like in practice

A causal map produced this way carries a complete audit trail. Pick any edge in the final map, click through, and you see the bundle of underlying claims, each with its source, its verbatim quote and its raw extracted labels. You can also see which transforms have been applied between the raw coding and the rendered view: which sources were filtered, which labels were rewritten, which evidence thresholds were imposed.

This is the property the special-issue call describes as "accountable links between analytic claims and empirical materials". It is not unique to causal mapping, but the structure of the workflow makes it easy to deliver. Every analytic claim in the final paper reduces to a sequence of explicit operations on a links table whose rows are quote-grounded extractions. A reader who doubts the analysis can rerun the pipeline, change a single step and observe the consequences. A coder who disagrees with a particular extracted link can reject it without disturbing the rest of the analysis.

The contrast with black-box synthesis is sharp. An AI-generated thematic summary is, in effect, a single opaque step from corpus to conclusion. The links-table workflow replaces that opaque step with a sequence of explicit, locally checkable ones, with the AI confined to the first and narrowest of them.

The same audit trail is what makes the map useful for exploration. Picking an edge and reading the underlying quotes is often where analysis begins rather than where it ends. The workflow does not aim to produce a finished diagram and stop. It aims to make a corpus tractable enough that the analyst can decide where to read closely, where to push back, and where to look for the cases that do not fit.

## 7. Ethical and practical transparency

The Text & Talk call asks authors to declare the type of AI application used, its role in the workflow, and how analytic accountability was maintained. The workflow described here gives clean answers to those three questions.

- **Type.** Commercial large language models are used at the extraction step. Specific model versions are reported, because output behaviour changes with model versions and a replication needs to know which version produced which links.
- **Role.** The model is used only to propose candidate links from short text chunks, each paired with a verbatim quote. It is not used to summarise across documents, to choose the codebook, to write the report, or to make any claim about the world.
- **Accountability.** Every link in the final analysis is traceable to a specific quote and source. Every analytic step beyond extraction is deterministic and human-authored. The links table is an artefact that can be shared with reviewers and co-researchers and inspected line by line.

There are practical ethical considerations beyond declaration. Where source material is sensitive (interview transcripts, clinical narratives, evaluation data from vulnerable groups), the choice of model matters, because some commercial services may train on inputs or store data in ways that conflict with consent and confidentiality commitments. We use settings that disable training on inputs where available, prefer providers with explicit data-handling commitments, and remove personally identifying information at the chunking stage where possible. Consent processes should make AI use explicit, including the kind of model, the role it plays and the safeguards applied.

There is also a softer accountability point. Because the prompt is the codebook in another form, sharing the prompt is sharing the method. We make our extraction prompts available as part of the analytic record. A reviewer can read the prompt and form a view about whether it framed the extraction task fairly.

## 8. Limits

The workflow does not solve every qualitative research problem, and we do not claim that it should.

- **Non-causal meaning.** Some valuable content in texts is not causal: identity work, norms, emotions, metaphors, turn-by-turn interactional structure. Causal coding ignores these, in the same way that conversation analysis ignores other features. For research questions about how participants account for what makes a difference, causal coding is on point. For research questions about how participants do identity work or manage stance in interaction, it is not.
- **Claims are not facts.** A coded link records what a source claimed. It does not establish what is true in the world. The map is a structured record of evidence, and the move from evidence to truth requires the same epistemic care it would in any qualitative study.
- **Frequency is not effect size.** Counts of sources and citations measure how widely a claim is made in the corpus.

They do not measure the magnitude of any underlying causal effect.

- **The transitivity trap.** If Source A says **Training** -> **Knowledge** and Source B says **Knowledge** -> **Adoption**, it is tempting to read a path from **Training** to **Adoption**. Unless within-source thread tracing is imposed and reported, that stitches together a mechanism that no one in the corpus actually claimed. Reasoning over paths is one of the things the links table makes easy, which means it is also one of the things to be careful with.
- **The AI does not eliminate coder bias.** It relocates the bias into the prompt. The advantage of the workflow is that the prompt is an explicit, shareable artefact whose effects can be inspected and revised, but it is still doing the work that a human codebook would otherwise do.
- **Attention is local.** Because the model sees one chunk at a time, it cannot pick up cross-references or implicit qualifications that span chunks. Expanding the context window helps but is slow and expensive. For research questions where cross-document context is the point, this workflow is not on its own enough.

## 9. Relation to neighbouring traditions

Causal QDA is compatible with thematic analysis and qualitative content analysis (Mayring 2000). It can be used as a front-end that surfaces a causal layer of meaning, with thematic work running alongside for non-causal features. It differs in that it stores ordered pairs and not isolated categories, which makes downstream pathway analysis possible.

In relation to conversation analysis and discourse analysis, causal QDA is doing something different. It is interested in the content of what participants say causes what, rather than in how that talk is constructed turn by turn. The two are complementary rather than competing. A single research project can perfectly well combine a fine-grained interactional analysis of a subset of episodes with a corpus-wide causal-mapping pass.

Narrative analysis is closer. Both approaches take participants' causal accounts seriously as objects of study in their own right, whether or not those accounts are independently verifiable. Causal QDA is one structured way of building a corpus-wide picture of the causal claims that narrative analysts read closely.

A different recent proposal for AI-assisted QDA replaces coding with structured conversation between the analyst and the model (Friese 2025). That is an interesting move for some kinds of interpretive work. The workflow we describe goes in the opposite direction: it preserves a strong intermediate representation, a quote-grounded links table, precisely so that the AI never holds the analytic state in its own working memory. We do not see the two approaches as exclusive, and a hybrid is plausible, but the trade-offs are different.

## 10. Conclusion

The AI question in qualitative work tends to be framed at the wrong level. Asked whether to trust a model with thematic synthesis or with the interpretation of a corpus, most qualitative researchers reasonably refuse. Asked whether to use a model for one narrow extraction job at which it is locally accurate and locally checkable, while every other analytic decision stays in human hands, the answer can be yes. The workflow described here is what that yes looks like in practice: an AI that stays within the narrow limits given it, a human who never delegates judgement, and a links table that anyone can read in the middle.

Causal mapping earns its modest place in that workflow because the unit of coding is small, the intermediate product is structured, and the chain from any analytic claim back to the supporting quote is short and inspectable (Ackermann & Maytorena-Sanchez 2024; Narayanan 2005). Those are also the properties that make AI assistance honest. The fuller methods treatment of causal mapping as a member of the QDA family is given separately (see [Causal mapping as causal QDA](#)); here we have used just enough of it to support the AI argument.

## References

- Ackermann, & Maytorena-Sanchez (2024). *Overlooked and Underused? The Benefits and Challenges of Using Causal Mapping for Project Studies*. <https://doi.org/10.1016/j.plas.2024.100161>.
- Axelrod (1976). *The Analysis of Cognitive Maps*. In *Structure of Decision : The Cognitive Maps of Political Elites*.
- Braun, & Clarke (2021). *Thematic Analysis : A Practical Guide*. SAGE Publications Ltd. <https://www.torrossa.com/it/resources/an/5282292>.
- Eden (1988). *Cognitive Mapping*. [https://doi.org/10.1016/0377-2217\(88\)90002-1](https://doi.org/10.1016/0377-2217(88)90002-1).
- Evaluation (2024). *Causal Mapping*. <https://www.betterevaluation.org/methods-approaches/methods/causal-mapping>.
- Friese (2025). *Conversational Analysis with AI - CA to the Power of AI: Rethinking Coding in Qualitative Analysis*. <https://doi.org/10.2139/ssrn.5232579>.
- Kim, & Andersen (2012). *Building Confidence in Causal Maps Generated from Purposive Text Data: Mapping Transcripts of the Federal Reserve*. <https://doi.org/10.1002/sdr.1480>.
- Mayring (2000). *Qualitative Content Analysis*. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research. <https://doi.org/10.17169/FQS-1.2.1089>.

Narayanan (2005). *Causal Mapping: An Historical Overview*. In *Causal Mapping for Research in Information Technology*.  
[https://www.google.co.uk/books/edition/\\_/61z36j6QgmAC?hl=en&gbpv=1](https://www.google.co.uk/books/edition/_/61z36j6QgmAC?hl=en&gbpv=1).

Powell, Copestake, & Remnant (2024). *Causal Mapping for Evaluators*. <https://doi.org/10.1177/13563890231196601>.

Powell, Cabral, & Mishan (2025). *A Workflow for Collecting and Understanding Stories at Scale, Supported by Artificial Intelligence*.  
SAGE PublicationsSage UK: London, England. <https://doi.org/10.1177/13563890251328640>.

Remnant, Copestake, Powell, & Channon (2025). *Qualitative Causal Mapping in Evaluations*. In *Handbook of Health Services Evaluation: Theories, Methods and Innovative Practices*. [https://doi.org/10.1007/978-3-031-87869-5\\_12](https://doi.org/10.1007/978-3-031-87869-5_12).

{wickham (2011). *The Split-Apply-Combine Strategy for Data Analysis*. <https://doi.org/10.18637/jss.v040.i01>.