



# WORKFLOWS - BARBARA

📅 29 May 2025

Yes, well you can do automated causal mapping coding inside [Causal Map](#), an online app which was originally designed for manual coding ("Nvivo but for stories"). We now have a new mega-app called [Workflows](#) which is specifically designed for causal mapping of large amounts of text in a reproducible/verifiable way – that isn't open for the public at the moment, we use it ourselves for clients' causal mapping projects (that is our business model).

I also really liked your comment to the presentation on how to quality assure / quality test AI outputs and discussed something similar with the presenter after the session. He suggested using the “reasoning model” of ChatGPT (which I don't know anything about as I'm not an AI expert) because AI is able to say “there's no simple answer to this and it depends on how you define “variety”” etc. ideally providing examples of different correct answers based on different definitions.

We are pretty opinionated about these things! In a nutshell:

We think:

- Giving any AI the freedom to make evaluative judgements as a black box is all wrong.
- vague "big black box" questions like "summarise this document" or "was this project effective" should be deconstructed into multiple "tiny black box" tasks like "does this paragraph mention health behaviour" which are much more intersubjectively verifiable / have good inter-rater reliability and which an AI can handle transparently.
- The best *generic* way of making sense of a large number of documents for evaluation purposes is to code each and every causal claim ("causal mapping") and collect those claims into a database of claims: a causal map or a causal knowledge graph. This can be quite easily done by an AI with thousands of "tiny black box" tasks.
- We see causal mapping in this sense as "causal QDA" which has some advantages over other kinds of QDA:
  - you don't have to think so hard about "what themes should I code for"
  - for evaluation purposes, causal information is 75% of what you need to gather

- Many evaluation questions can be then (at least partially) answered more or less automatically by querying that causal map, e.g. *show me all the causal pathways from Intervention X to Outcomes Y and Z which were each mentioned in their entirety by the same individual or report, ignoring the claims which were assessed as dubious.*
- The result of all this is essentially a set of instructions, a script or algorithm (nothing essentially to do with AI, just a long, nested list of tasks which a tireless human could carry out) which deconstructs high-level evaluation questions into a series of intersubjectively verifiable tasks. Our Workflows app helps the user create that script, which you then basically just run, wait as documents get coded and tables and maps appear in answer to your questions. You can run it again the next day or with a similar but different AI model or a bunch of humans and you should get similar answers. The human is very much in the loop in the sense of designing this script, but it's high-level input, we can then go and watch the flowers grow while the AI does the grunt work.
- Caveats
  - This is all about volume and scale and making sense of lots of different sources or documents containing fairly generic causal claims, e.g. from diverse stakeholders often in multiple languages. It's kind of the opposite of process tracing and its intricate, high-stakes examination of a fairly small number of detailed causal claims (I think).
  - Causal mapping is not really a method of causal inference: all it does is assemble evidence about what these sources think, it's still up to the evaluator to decide what conclusions can be drawn from that. (QuIP on the other hand, though it uses causal mapping, is more explicit about how to make causal conclusions from causal mapping data.)

re the "reasoning model" the presenter mentioned - we think that is exactly the wrong way round: give loads of evaluative freedom to a big black box and then also ask it to invent a plausible story about how it reached that conclusion. To be fair, some of the newer reasoning models really are able to display their actual working retrospectively. That's quite an improvement because it offers some retrospective transparency (a bit like saying to an evaluator, don't submit an inception report or an evaluation matrix in advance, just do your magic and then when you present your conclusions tell us how you arrived at them). I still think it is the wrong way round though.