The point of us – when AI can do the desk analysis, the learning between people is the output

31 Oct 2026

For evaluators, commissioners and evaluation managers who are asking what an evaluation is for, now that AI can do much of the analysis.

Work in progress. A position paper. Most of it argues from our own practice and from published methods; where we speculate, we say so.

In short

- **The desk work is going.** In our experience, a frontier AI model given an evaluation's terms of reference and documents can now answer the desk-based part of many evaluation questions about as well as a competent evaluator, and the models improve every few months.
- **That raises the question a friend put to us: what is the point of us?**
- **The report was never the only output.** Evaluations also change the people who take part in them: what they believe causes what, whom they know and trust, what they notice, and how ready they are to act. Evaluators have a name for part of this, process use (Patton 2008).
- **We have mostly treated those changes as a by-product**, or as a necessary and costly step on the way to the report. While the desk analysis took weeks, it competed with them for the budget, and it usually won.
- **When the desk analysis takes minutes, the balance changes.** For many evaluations, the learning that happens between people becomes the main thing an evaluator can add.
- **So design for it.** Decide whose learning matters, involve those people in the work itself rather than sending them messages about it, make them work with each other, and look for evidence that something changed.
- **AI can help**, by drafting, synthesising and widening who can take part. It cannot take part on anybody's behalf.

See also: What is the Point of Us. A Sci-Fi Story for Researchers; Pilots, not passengers — reconstructible reporting in evaluation; How hard is evaluation actually.

The desk work is going

We have been testing what frontier models can do with an evaluation's documents. Given the terms of reference and a folder of transcripts and reports, they can code, compare groups, test a theory of change link by link and draft findings, in minutes. We built Rubicon on the assumption that they will keep getting better at it.

What has not become easier is telling. Earlier we wrote that selecting and applying a method, which we used to think of as the hard part, suddenly looks easier, "Whereas the stuff which has been neglected, like establishing a rapport, knowing which question to ask and when, or reading an undercurrent, does not look very much easier" (How hard is evaluation, actually?). Those are all things that happen between people.

What is the point of us?

A friend put this to one of us during the 2025 argument about generative AI in qualitative research. If AI can do the technical work and then make meaning better than we can, what is left? We answered with a short science-fiction story about a woman who writes a thesis that an AI could have written better, and writes it anyway, because "The point of the journey is the journey, and how it changes her and her fellow humans."

An evaluation is not a private thesis. It has a commissioner, questions somebody needs answered and a deadline. But the story's answer carries over further than it first seems to, because evaluations have always changed people, and the people they change are often the ones who decide what happens next.

What evaluations produce besides the report

Besides the report, an evaluation can change the people who take part in it:

- **Changed causal beliefs.** A programme manager who has sat through the coding of twenty interviews may come away believing different things about what drives the outcomes from one who read the summary. Causal mapping exists to capture people's mental models of how change happens; the same tools can show whether the models of the people who took part changed over time.
- **New relationships.** Frontline staff and managers, commissioners and implementers, communities and funders, arguing over the same evidence, sometimes for the first time.
- **Trust in the findings**, because people saw how the findings were made and had their chance to object.
- **Noticing.** Participants start to see things in their own work that they did not see before.

- **Commitment to act** on what was found, which comes more easily to people who helped to find it.

None of this is new. Patton called the changes in thinking and behaviour that come from taking part in an evaluation "process use" (Patton 2008). Some methods put it at their centre. The Most Significant Change technique is built around stakeholders discussing which stories matter most, and why (Davies & Dart 2005). In outcome harvesting, the harvest workshop and the drafting loop are conversations between people; in one published harvest, some of the organisations involved spent two to four times as long on it as the evaluators did, and as we put it in our outcome harvesting paper, "In a participatory method the iteration is the method." Group model building aims at a model of the problem that participants share and are committed to (Vennix 1999). Guijt argued for monitoring designed around collective learning (Guijt 2008).

Our own Garden has looked past it. In The result of an evaluation is a qualitative causal model we asked what the output of an evaluation is, answered "a report", and added: "In the sense of developmental evaluation perhaps some learning has taken place as well, or as a main output, but this is not what I want to address here." This page addresses it.

A by-product, or an annoying extra step

In the evaluations we have worked on, the learning between people was usually treated in one of two ways.

- **As a by-product**: welcome when it happened, not planned for, not budgeted and not reported.
- **As a step on the way to the report**: a validation workshop to check the findings, or consultation to get better data or buy-in, so that the report would be believed and used. Necessary, often valued, and the first thing cut when the days ran short.

What came after the report was usually messaging: a summary, a slide deck, a webinar, a set of key messages. Messaging tells people what was found. It rarely changes what they believe about how things work, because they did not take part in finding it. We notice the same thing in ourselves. Reading an analysis an AI has written, "we sometimes still struggle to care. Perhaps because we didn't get our hands dirty enough writing the report, or don't have enough skin in the game" (Saturation, scale and attention). If that is true of evaluators reading a machine's report, it is true of stakeholders reading ours.

When the analysis is cheap

While the desk analysis took most of an evaluation's days, participation competed with it for the budget. A team that spent three weeks coding had little left for a second workshop. Now that a first analysis takes minutes, the days go somewhere else, and the question is where.

Our answer is that they should go to the people. Not every evaluation should spend them this way, but in many evaluations the changes in the people involved are what make a difference afterwards, and they are the part of the work a machine cannot do for us. An evaluator who can get the right people into the room, involve them in the analysis and help them work through disagreement is doing work that does not get easier as the models improve.

AI also changes who can take part in the analysis. Coding and querying a whole dataset used to need training and software. A community group, a team of frontline staff or a board can now put their own questions to the evidence, in their own language, and compare what they find with what the evaluator found. Analysis was once something done for stakeholders; it can now be something done with them, or by them.

Speculation: we expect some commissioners to start asking for this outright: for an evaluation's terms of reference to say whose understanding it should change, and for the report to include a section on what changed among the people who took part.

Designing for learning between people

If the learning between people is an output, it deserves the planning we give to a sample or an analysis. We suggest six principles.

- **Start from whose learning matters.** At inception, ask whose understanding, or which relationships, would need to change for the evaluation to make a difference. Plan for those people as carefully as for respondents.
- **Involve them in the work, not just the messaging.** The work includes choosing the question, agreeing what counts as good enough, coding a sample, arguing with a draft finding and deciding what is significant.
- With AI doing the first pass, each of these can take an hour rather than a week, so more of them fit into an evaluation.
- **Make them work with each other, not only with the evaluator.** Mix the tables. Pair people who rarely talk. Give a dissenting view a seat and time.
- **Allow more than one round.** People change their minds between meetings, not during them. The drafting loop in outcome harvesting is a model.
- **Respect their time.** Participants' time is not free, and most of them are not paid to evaluate. Say whose time a design spends, and spend it on the steps where it changes most.
- **Look for evidence that something changed.** If learning is an output, evaluate it. A short interview with each participant at the start and the end, asking what they think drives the outcome, can be coded into causal maps and compared. A change that cannot be seen in any record should not be claimed.

AI can help with most of these, in four ways.

- **More voices.** AI interviews let many more people contribute, on their phones and in their own language, than a team could interview in person.
- **Faster synthesis.** In a workshop, an AI can summarise what each table said and set the tables side by side, so the room spends its time on the differences.
- **Something to argue with.** A machine-drafted table of findings gives a workshop a starting point, as long as it is marked as a draft and the room decides what stays.
- **Time.** A first analysis in minutes frees days for meetings.

What AI cannot do is take part on anybody's behalf. A programme manager does not change their mind because a model has read the interviews.

What this does not mean

- **It does not mean the analysis stops mattering.** A well-run process built on a wrong analysis does harm, and it does it persuasively. The desk analysis still has to be right and open to checking, which is what Rubicon is for.
- **Not every evaluation should work this way.** Some accountability evaluations need distance from the people being evaluated, and close involvement can slide into capture.
- **Participation can be tokenistic, and power decides who is in the room.** "Participation" covers everything from being informed to being in control (Cornwall 2008). A design that involves the people already in charge, and messages everybody else, is the old design with more meetings.
- **Learning is hard to evidence, and easy to overclaim.** Before-and-after interviews help, but people may tell an evaluator what they think the evaluator wants to hear.
- **Our claim about what AI can do comes from our own tests and practice.** It is not yet established across the field, and it may be truer of some kinds of evaluation question than others.

I sometimes use AI to help draft these pages, but the final text is ours.

References

Cornwall (2008). *Unpacking `Participation': Models, Meanings and Practices*. Community Development Journal, *43*, 269--283. https://doi.org/10.1093/cdj/bsn010.

Davies, & Dart (2005). *The Most Significant Change Technique, A Guide to Its Use*.

Guijt (2008). *Seeking Surprise: Rethinking Monitoring for Collective Learning in Rural Resource Management*. publisher not identified.

Patton (2008). *Utilization-Focused Evaluation*. Sage.

Vennix (1999). *Group Model-Building: Tackling Messy Problems*. System Dynamics Review, *15*, 379--401. https://doi.org/10.1002/(SICI)1099-1727(199924)15:4<379::AID-SDR179>3.0.CO;2-E.