

ESSAY 07

BETTER JUDGMENT IN THE AGE OF AI

A Human in the Loop Is Not the Same as Human Judgment

"The presence of a human does not prove the presence of judgment."

A Human in the Loop Is Not the Same as Human Judgment

By Alen Mayer

Somewhere along the way, "there's a human in the loop" became one of the standard answers to questions about AI oversight.

Ask whether a hiring system is safe, whether a lending decision is fair, whether an automated recommendation gets checked before it does damage — the reassurance arrives in the same four words. A human is in the loop.

It's worth asking what that sentence actually guarantees.

One useful framework argues that organizations need to deliberately decide where humans sit relative to an AI system — in the loop, on it, or out of it entirely — especially for consequential decisions. That's a real design choice, and making it on purpose beats leaving it to accident. But deciding where a human sits answers a structural question. It doesn't answer the harder one: once she's there, what is she actually doing?

This isn't a question about whether anyone is watching. It's a question about what watching requires — a different failure than an empty seat, and a quieter one, because from the outside they look identical.

■ APPROVED IN FOUR SECONDS

An AI system evaluates four hundred job applications and recommends thirty for interviews. A hiring manager reviews the thirty and approves twenty-five.

On paper, that's exactly the design working as intended. Human oversight, present and accounted for. Technically, human review. ✓

But what did the review actually consist of? Did she read the reasoning behind each recommendation, or scan a name and a score? Did she look at any of the three hundred seventy the system didn't surface, to see whether the cutoff made sense? Did she understand what the system was using to make the recommendation, well enough to question whether it deserved her approval? Or did she spend four seconds on each of the thirty, because the recommendations looked reasonable and there was no obvious reason to slow down?

Both versions of that manager produce an identical record. Same timestamp, same approval, same audit trail showing a human in the loop. One of them exercised judgment. The other exercised a click.

■ TWO DIFFERENT THINGS THAT LOOK THE SAME FROM OUTSIDE

Human review is a process condition — a fact about where a person sits in a workflow, checkable by looking at an org chart or a sign-off field. Human judgment is a cognitive act — understanding the recommendation, weighing the evidence, looking for what may be missing, and deciding whether the recommendation deserves action. The workflow can require the review. It cannot guarantee the judgment.

The workflow can require the review. It cannot **guarantee** the judgment.

An organization can build a perfectly designed loop, put a genuinely qualified person in it, and still get nothing but the process condition, because the judgment depends on things the structure can't compel: whether she has time to actually look, whether she has permission to actually disagree, whether disagreeing has ever once been rewarded rather than treated as friction, whether she was ever given a reason to believe her second look might catch something the system missed.

None of that shows up in the design. All of it determines whether the loop is doing anything.

■ WHAT ACTUAL JUDGMENT REQUIRES

Meaningful human judgment inside a loop isn't a mystical extra ingredient. It requires specific conditions and behaviors, and many reviews quietly skip several of them.

It requires understanding what the system concluded and roughly why, not just what score it produced. It requires knowing which piece of evidence the recommendation actually rests on, so a person can ask whether that evidence deserves the weight it's been given. It requires noticing what's absent — the applicant who didn't make the list, the case the system never flagged — since judgment that only inspects what was surfaced can never catch what wasn't. It requires a genuine willingness to disagree, which is different from having the formal right to; a lot of organizations grant the right and quietly punish the exercise of it. And it requires actual authority to act on that disagreement, because a reviewer who can object but not change the outcome isn't exercising judgment. She's filing a comment.

Skip enough of that list and what remains is a person standing where judgment is supposed to happen, performing none of it.

■ THE MISTAKE THIS ENABLES

The dangerous version of this isn't a company deliberately faking oversight. It's a company that genuinely built the loop, believes the loop is working, and stops looking

further — because the box that says "human reviewed" is checked, and checking that box was the whole design goal.

That's the trap worth naming. A human in the loop was never meant to be the finish line. It was meant to be the place where judgment happens. When the loop exists but the judgment inside it quietly doesn't, the organization gets all of the reassurance and none of the protection — and has no way of knowing that, because the paperwork looks exactly the same either way.

The presence of a human does not prove the presence of judgment.

TAKE THIS FURTHER

Build better judgment into how your team works with **AI**.

I help leaders and organizations strengthen the human capabilities behind better judgment as AI becomes part of everyday work and decision-making.

JUDGMENT LABS

A 90-minute session working through realistic AI-influenced business decisions.

WORKSHOPS

Practical training on how teams question, interpret, weigh and decide with AI.

KEYNOTES

Reframing what human judgment requires as AI becomes more capable.

Alen Mayer

alenmayer.com