ESSAY 02

BETTER JUDGMENT IN THE AGE OF AI

When AI Is Right and the Decision Is Still Hard

"The hardest AI judgment problems don't begin when AI fails. They begin when AI works exactly as designed, and a human still has to decide what follows."

Alen Mayer

alenmayer.com

When AI Is Right and the Decision Is Still Hard

By Alen Mayer

For most of my career, a bad decision usually gave you somewhere to look.

Someone skipped a step. Someone worked from incomplete information. Someone trusted an instinct that didn't hold up once you looked closely. Find the gap, and you'd often found the problem.

AI is creating a more difficult category of decision.

Nothing has to break down.

The information can be right. The recommendation can follow logically from it. The system can do exactly what it was built to do.

And the decision can still be genuinely hard.

We have spent enormous energy asking what happens when AI is wrong. The next judgment problem is what happens when AI is right.

FOUR ROOMS, NONE OF THEM ABOUT AI BEING WRONG

A sales team's AI ranks accounts by expected revenue. It has the full account record — it knows this particular customer sits on the product advisory board, has been a reference account for years, appears in the case studies on the company's own website. The ranking still comes back low, because revenue is what it was built to optimize, and nothing about advisory boards or case studies was ever part of that. The AI didn't miss anything. It simply wasn't told that kind of thing should matter.

An HR team retires a productivity metric that ignored case difficulty, replacing it with one that weights each case by how hard it's historically been to resolve. It works exactly as designed. Then someone notices that the new, more accurate score says nothing about how much of that performance came from the employee and how much came from the AI assistant. That may not matter if you're measuring productive output in the role as it's currently tooled. It matters considerably more if you're using the same number to judge individual capability for promotion. The metric isn't wrong. The question has changed.

A pricing agent has approved thousands of small deals with a 98% clean-outcome rate — no disputes, no clawbacks. Leadership wants to triple what it's allowed to approve on its own. But even if the same 98% held at the higher level, that wouldn't by itself settle the authority question. The probability might stay exactly the same while the consequence of each mistake becomes much larger.

An operations system reallocates production capacity exactly the way leadership designed it to — optimizing margin and delivery, the way everyone agreed it should. This time, the plant it deprioritizes happens to be mid-way through a strategic expansion nobody built into the formula, on purpose, because giving every plant a "strategic" exception was exactly the kind of political drift the system was built to end. An executive doesn't like the result. The system didn't malfunction. The rule did precisely what it was designed to do — and now someone has to decide whether that design still holds up.

WHAT GETS MISSED WHEN WE ONLY LOOK FOR FAILURE

None of those four situations requires the AI to have failed. The data can be correct, the system can execute as designed, and no policy needs to have been violated. If your only question is "did the AI get this right," the answer can still be yes — and that answer stops being useful the moment you ask it, because being right was never actually where the difficulty lived.

That's the trap. Verifying AI output has become, correctly, a basic discipline. But it quietly teaches people that once something checks out, the hard part is over. In each of these rooms, checking out is exactly where the hard part began.

THE FOUR QUESTIONS THAT SURVIVE A WORKING SYSTEM

Underneath those four situations are four distinctions that correctness alone cannot resolve:

Accurate information is not the same as a correct objective.
An accurate metric is not automatically evidence for the decision you're using it to make.
A reliable track record does not, by itself, justify a bigger grant of authority.
Correct execution inside a rule does not settle whether the rule deserves your continued confidence.

None of those questions goes away as the underlying AI gets better. AI may help enormously with each one — it can analyze objectives, test whether a metric fits a decision, model consequences, even recommend where an authority boundary should sit. But the fact that its output is correct doesn't determine the answer to any

of them. If anything, better AI makes the questions more consequential, because better AI is exactly what makes organizations comfortable acting faster on what it produces. A system you don't quite trust gets checked by habit. A system that's earned your trust gets deferred to — and deference is precisely the condition under which an unexamined objective, an ill-fitting metric, an over-extended authority, or a rule nobody's revisited can do the most damage before anyone notices.

> **The hardest AI judgment problems don't begin when AI fails. They begin when AI works exactly as designed, and a human still has to decide what follows.**

THE PRACTICE THAT ACTUALLY HELPS

I don't think the fix is treating every AI-assisted decision with fresh suspicion — that's exhausting, and it isn't what any of these four situations actually needed. What each of them needed was a single habit: before accepting a correct output as a closed question, ask which of the four questions it's quietly resting on. Is this about what we're optimizing for? Is this metric evidence for what I'm about to decide, or just accurate about something else? Does being reliable here mean it deserves more authority there? And if I designed this rule myself, would I still approve it, now that I've seen what it costs?

That's not a checklist to run on every decision. It's a reflex to build for the ones that matter — the pipeline review, the promotion cycle, the authority you're about to expand, the outcome that made you wince. Most decisions won't need it. The ones that do are exactly the ones nobody thinks to ask, because everything already checked out.

Better AI doesn't make judgment optional. It moves the judgment upstream — into the objectives, the evidence, the authority, and the rules nobody has looked at since the day they were approved.

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