

ESSAY 13

BETTER JUDGMENT IN THE AGE OF AI

The Tool Wasn't Wrong. We Asked It to **Decide** Too Much.

"The tool didn't fail. The decision it had become part of was never made on purpose."

The Tool Wasn't Wrong. We Asked It to Decide Too Much.

By Alen Mayer

A sales team starts using AI to summarize customer calls. The summaries are accurate, consistent and fast. Nobody objects.

Then something happens that nobody decides.

■ THREE STEPS NOBODY APPROVED

Managers start using the summaries to judge how healthy a deal is. Forecasts begin to lean on those health assessments. Eventually, a rep's performance is measured against the forecast.

Call summary → deal health → forecast → performance evaluation

Each step is small and reasonable. Each is easy to defend on its own. But follow the path from start to finish and a tool that was introduced to summarize calls is now influencing how people are evaluated.

At no point did anyone decide that an AI call summary should shape a performance review.

■ ACCURACY DOESN'T TRAVEL

A summary can be excellent at what it was built to do: capturing what was said. It tells you much less about what the buyer meant, how serious they are, or whether the pause before the pricing question was hesitation or just a bad connection.

That gap doesn't matter while the summary is used to remember the call. It matters a great deal once the summary is used to judge the deal.

But the tool arrives at the new job carrying the credibility it earned at the old one. It was accurate before, so it feels reliable now.

But reliability does not automatically transfer across purposes. A tool can be valid for summarizing a conversation and untested for judging deal quality.

A tool earns **trust** for a task. It does not earn it for the next one.

■ WHY IT'S HARD TO SEE

When people review AI output, they usually ask whether it's accurate. In this case, the answer can remain yes even while the use keeps expanding.

The question that goes unasked is a different one: what decision is this now informing?

That question isn't a technical check. Nobody can answer it by examining the tool. It can only be answered by looking at how the output is actually being used.

■ WHAT CHANGES THE RISK

This isn't an argument against using AI summaries, or against putting them to wider use. Sometimes the wider use is justified. The point is that it should be a decision, made on purpose, about a new use of the tool, and not something that happens because the output was already available.

The biggest risk may not be that a tool fails at what it was designed to do. It may be that we start relying on it for decisions it was never evaluated to support.

■ WHAT LEADERS CAN DO

For any AI output that people rely on, ask two questions:

"What decision does this now influence?"

"Was it ever tested for that?"

If the decision it influences has changed, treat the use as new, even if the tool has not. Look at it the way you would if someone proposed it for the first time.

The tool didn't fail. The decision it had become part of was never made on purpose.

Companion Judgment Challenge: "The Summary That Became a Scorecard."

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