WORKBIGLY

AI · A BRIEF FOR LEADERSHIP

Picking the tool is the easy part. Deciding where judgment sits is the job.

Four decisions no one else can make for you, and the few definitions you need to make them.

THE STARTING POINT

It shows up capable.
It doesn't show up on your side.

A model turns up knowing an enormous amount, and none of it is what your business should do next. It has never seen your operation, your customers, or the call you made last year, and nothing bad happens to it when it gets something wrong.

That isn't a defect. It's the shape of the product. The capability is solved and shipped; the judgment isn't, and no vendor can sell you that part, because it runs on things only your own people know.

Which is why this looks less like a purchase and more like four decisions. All four sit with you, and not one of them is technical.

The one thing not in the box is judgment

And it's what decides whether two companies buying the same tool end up in the same place.

DECISION 1

Where to put it: where time runs short, not judgment

Working through free text at volume. Sorting things into a structure the company already defined. Recasting one piece of content for a different audience or a different language. Keeping up something repetitive without getting sloppy.

The four have one thing in common, and it's what makes them a safe place to begin: what runs short is time, not judgment. That's where the gains have actually been measured — 14% more cases resolved per hour on a support desk, 26% more tasks finished by developers — and none of it asks the organization to change anything yet.

Read the same test the other way and you have the limit. Where judgment is what's short, the tool does not supply it. If the first use lands there, who supplies the judgment the tool doesn't bring?

Brynjolfsson, Li and Raymond, NBER w31161 · Cui et al., Management Science 2025

DECISION 2

How much it decides on its own

An assistant answers, and your people choose what happens next. A workflow runs steps your people laid out in advance. An agent picks its own next move on the way to a goal.

All three get sold under the same word. What actually changes is how much of your team's judgment stays inside the work. With an assistant, all of it stays. With a workflow, it lives in the design. With an agent, it sits outside the path entirely — nowhere except in the limits you set up front.

For most of what a business actually runs on, a well-defined workflow does better and fails in ways you can predict. Reaching for the most autonomous option by default pulls judgment out of somewhere it was working, and charges you for the trouble.

DECISION 3

Who checks it, and against what

A search engine fails by handing you a bad link, and you notice. A model fails by writing a clean, confident paragraph that happens to be untrue. That's a hallucination, and it isn't the fault of whoever asked: nothing in the writing gives it away.

Since the failure is invisible, checking can't be left to whoever happens to spot it. Somebody has to decide, in advance, which outputs get checked, by whom, and what they get checked against. Against the source. Against a procedure. Against the judgment of a named person. Never against "common sense" — that's trust, not a control.

And make it a role, not a rule. You can assign a role, train it, and give someone credit for holding it. "Everyone review everything" allows none of the three, and nobody does it anyway.

McKinsey 2025 · on defined processes for human validation

DECISION 4

What gets written down

When the process is clear, the tool makes it faster. When the process is muddled, the tool makes it faster and muddier. Automating something nobody ever wrote down doesn't clear it up; it fixes the ambiguity in place and runs it at scale.

So there's a cheaper, duller decision that comes before the tool: which processes get written down, and what one person currently carries in their head that ought to be on paper. Why that machine gets purged before startup. Why that customer gets invoiced differently.

That one pays off with or without AI, and it's the only one of the four still worth anything the day you change tools.

BEFORE YOU SIGN

Five questions, and not one of them technical

Which number in my business does this move? What happens when it gets something wrong — when, not if — how often, and who catches it?

Where does anything we put into it end up, and who can get at it? What record is kept of how it reached what it says? And what is it allowed to do without anyone signing off?

All five are answered with evidence from your own operation, not with a demo. And it's worth knowing that close to half of organizations still haven't settled when an output needs a person to look at it.

SOURCES

Where each number comes from

14% more cases resolved per hour, across 5,179 customer-support agents: Brynjolfsson, Li and Raymond, "Generative AI at Work" (NBER w31161).

26% more tasks finished: Cui et al., Management Science (2025), three field experiments covering 4,867 developers at Microsoft, Accenture and a Fortune 100 company.

Close to half with nothing settled: McKinsey (2025) finds that 54% of organizations say they have defined processes for when an output needs validation.

All four fit in one meeting

Where to put it, how much it decides on its own, who checks it and against what, and what gets written down. Settle those and the tool is a line item.

workbigly.com