WORKBIGLY

AI AT WORK · A PLAIN-ENGLISH GUIDE

Everyone gets the same AI. Not everyone has the same data to feed it.

Where it's already paying off, what the words mean, and what separates two companies running the same tool.

WHERE IT'S ALREADY PAYING OFF

Measure a specific task and the numbers show up

In one trial, 5,179 support agents resolved 14% more issues per hour — 34% more among the least experienced. Across three trials, 4,867 developers completed 26% more tasks.

Meanwhile, a survey of nearly 9,000 firms found more than eight in ten executives saw no productivity impact at their own company over three years. Both are measured, and they don't contradict each other: one looks at the task, the other at the whole business.

The gap between them is organizational work. It's the only part of this you actually control, and it's where companies pull ahead.

Brynjolfsson, Li and Raymond, NBER w31161 · Cui et al., Management Science 2025 · NBER w34836. Full detail on the sources page.

Everyone can buy the same model. Nobody can buy your data.

The tool is there for everyone. What you've written down and structured isn't.

DEFINITION 1

A search engine and a model fail differently

A search engine tells you where the answer is. When it fails, you get a bad link and you know in two seconds.

A language model writes the answer for you. When it fails, you get a flawless paragraph that happens to be wrong. Nothing in the text tips you off — it reads exactly as confident either way.

That's the whole conversation about control right there. Neither one is worse. One fails loudly and the other fails quietly.

Hallucination: a false statement delivered with total confidence

It's not your fault for asking badly, and asking better won't fix it.

DEFINITION 2

Assistant, workflow and agent: three things, three price tags

An assistant answers what you ask, and you pick what happens next. A workflow runs steps somebody laid out in advance. An agent decides on its own what to do next to hit a goal.

All three get sold as "agents." That's not a marketing quibble — it changes where the control has to sit.

When an assistant gets it wrong, it hands you bad text and somebody reads it first. When an agent gets it wrong, it's already done three things. Control moves out of the output and into the guardrails.

DEFINITION 3

A prompt is a request, not a spell

Clear instructions get better results, the same way a well-written request to a vendor does. That's as far as it goes.

No instruction, however well phrased, makes the system know your company. If the fact isn't written down anywhere, there's no clever way to ask for it.

So the jump doesn't come from learning to ask better. It comes from writing down what today lives only in somebody's head.

WHAT TO USE IT FOR

Four things that already work, no exotic purchase required

Making sense of free text at volume — open survey answers, incident reports, email. The stuff that gets skimmed and filed today.

Sorting things into a structure you already own. Not inventing categories: picking from the ones your company already defined.

Reworking the same content for a different audience or a different language, which is manual today and therefore doesn't happen.

Grinding through repetition without drifting, which is exactly where people do worst.

Sorting isn't inventing: it picks a branch that already exists

That's why you can check it in seconds, undo it, and put a number on how often it's wrong.

THE UNCOMFORTABLE PART

If it can't answer about your company, you found something

When a system working off your own documentation can't answer a question about the process, check the documentation before the tool: is that information written down anywhere?

Example: someone with twenty years on the floor can tell you in two seconds why that machine gets purged before start-up, and no document says it.

If it isn't written down, that knowledge retires when they do. The question the system couldn't answer is a free inventory of what your company knows and has never written down.

WHAT IT CAN'T DO

And the limits, without the hedging

It doesn't know your operation, or what happened last year, unless you tell it.

It doesn't replace a measurement or a walk on the floor.

It can state something false with total confidence, and nothing in the text gives it away. Anything that's actually going to be used gets read by a person first.

BEFORE YOU SIGN

Five questions that keep a vendor meeting honest

Which number of ours does the improvement get measured against? What happens when it's wrong — not if it's wrong — how often, and who catches it?

Where does our data live, and who can get to it? What gets logged about how it reached its answer?

And the awkward one: what can it do without anyone signing off? All five get answered with evidence from your operation, not with a demo. A demo shows you the best day it ever had.

SOURCES

Where every number on these pages comes from

+14% issues resolved per hour, +34% among the least experienced agents: Brynjolfsson, Li and Raymond, "Generative AI at Work" (NBER w31161), across 5,179 customer-support agents.

+26% completed tasks: Cui et al., Management Science (2025) — three field experiments with 4,867 developers at Microsoft, Accenture and a Fortune 100 firm.

More than eight in ten with no productivity impact at their own firm over three years: NBER w34836, "Firm Data on AI" (Revised October 2026), a survey of nearly 9,000 firms in seven countries.

Measure one thing, change one thing

One specific task, the number you write down before you start, and the date you'll go back and check it.

workbigly.com