

BEYOND FASTER ANSWERS

Faster Outputs, Same Decisions: The Productivity Paradox of Enterprise AI

Enterprise AI has made producing answers almost free — and left the act of deciding exactly as hard as it has always been.

Giovanni Leonardi

February 2022



Faster Outputs, Same Decisions: The Productivity Paradox of Enterprise AI

Transformation · No. 274

Originally written February 2022 | Re-edited and re-rendered for the WhereBeyond Edition, 2025

"A gap that is never funded is not a delay; it is a permanent state wearing the costume of a delay."

1. The forty-minute report that changed nothing

Picture a large programme in its Monday review. Over the past year the analytics team has cut the time it takes to produce the weekly performance pack from three days to about forty minutes. The pipelines are automated, the models retrain themselves overnight, the dashboards refresh before anyone is awake. Everyone in the room agrees this is progress, and they are right to. And yet the decision the pack exists to inform — whether to keep funding a workstream that has missed its milestones for two quarters — is the same decision the room deferred last month, and the month before that. The information now arrives faster than anyone thought possible. The choosing has not moved an inch.

This is the paradox I want to name, because we are living inside it and mostly failing to see it. For the last several years the enterprise has poured money into what we loosely call AI — meaning, in practice, machine learning, predictive analytics, and automation — on a premise that sounds unarguable: that better, faster information produces better, faster decisions. The premise is exactly half right. We have made the supply of answers astonishingly cheap. We have done almost nothing about the demand side — the organisation's capacity to take an answer and act on it. The result is the pattern in that Monday room, repeated across the economy. Faster outputs, same decisions.

2. We solved the cheap half

It is worth being honest about why the supply side fell first. Producing an analysis quickly is a well-posed technical problem: it has a clear input, a clear output, and a metric you can drive down. Deciding differently is not a technical problem at all. So the money flowed, entirely rationally, to the part that engineering could solve, and the hard part was quietly relabelled "adoption" and left to look after itself.

Economists have been circling this for a long time. Robert Solow's remark from 1987 — that you could see the computer age everywhere except in the productivity statistics — has aged into something close to a law. The more recent work on the productivity

paradox of artificial intelligence makes the same point with sharper edges: the technology can be genuinely transformative and the measured gains can still be absent for years, because the value does not live in the model. It lives in the reorganisation around the model, and that reorganisation is slow, political, and expensive.

Let me make this concrete, because abstraction is how this argument usually dies. Consider a demand-planning function that invests in a better forecasting model and genuinely improves it — say, cutting the average forecast error from a MAPE of twenty-two per cent to fourteen. That is a real gain, hard-won, the kind of number a data-science team can be proud of. And then watch what happens in the sales and operations planning meeting: the improved forecast is overridden, as it always was, by the number the sales director has committed to the board. The model got materially better. The decision that consumed it did not change by a single unit. Why would it? The binding constraint was never forecast accuracy. It was that a commit is a promise, not a prediction, and no amount of accuracy dissolves the incentive to defend a promise.

That is the mechanism the whole paradox turns on. We keep improving the accuracy and the speed of the inputs to decisions whose real constraint sits somewhere the model cannot reach.

3. "It is only a lag" — and why that is no comfort

The strongest objection to all this deserves to be stated at full strength, not as a straw man. It goes like this: You are not describing a paradox, you are describing a lag. The productivity J-curve predicts exactly this — the gains show up only after the complementary investments in process, structure, and skills catch up. Be patient. The decisions will improve once the technology beds in.

I have real sympathy for this, and it is partly true. The lag is real and the J-curve is a good description of it. But notice what the argument quietly concedes. It says the returns depend entirely on the complementary investments — the rewiring of process and decision rights — and not on the technology itself. That is not a rebuttal of the paradox. That is the paradox, restated by its defenders.

And here is where sympathy runs out. The J-curve only bends upward if someone actually makes the complementary investment. What I observe is the opposite: firms are enthusiastically funding the cheap half and systematically starving the expensive half, because buying another model is a procurement decision and rewiring who is allowed to decide is a fight. Left to itself the lag does not close. It calcifies. A gap that is never funded is not a delay; it is a permanent state wearing the costume of a delay.

A lag closes only if someone pays to close it. When the entire complementary investment is filed under "change management" and funded last, the J-curve does not bend — it flattens into a plateau, and we call the plateau maturity.

4. What actually holds a decision still

If speed of information is not the constraint, what is? In my experience three things hold a decision still, and no model yet built touches any of them.

- **Authority.** Whether the person who sees the answer is the person permitted to act on it. An insight delivered to someone without the mandate to move is not a decision aid; it is a source of frustration with a nightly refresh.
- **Consequence.** Whether deciding differently is safe. In most organisations the career cost of reversing a course you championed is far higher than the cost of quietly letting it run. When continuing is safe and stopping is exposed, the data can say stop as loudly as it likes.
- **Commitment.** Whether the organisation can hold a decision once it is made, or whether it re-litigates the same question every reporting cycle. This is the one that faster outputs can actively make worse: an inexhaustible supply of fresh dashboards gives everyone a fresh pretext to reopen a settled question. More information, more re-argument, less resolution.

A decision sits at the intersection of these three. Accelerating the production of answers pushes on none of them, and occasionally pushes the wrong way. This is why the forty-minute report changed nothing: the report was never the bottleneck. The bottleneck was a workstream nobody had the authority, the safety, or the collective will to kill.

5. Measure the decision, not the output

The corrective is not to slow the tools down; the tools are the good news. It is to shift the thing we measure. We have become fluent at instrumenting the latency of production — how fast the pack is built — and almost illiterate at instrumenting the latency of action: the time between an insight becoming available and the organisation actually doing something different because of it. That second number is the one that predicts value, and almost nobody tracks it.

So before signing off the next model, the next pipeline, the next self-service tool, it is worth forcing three questions and refusing to proceed without answers:

1. What specific decision will this change?
2. Who owns that decision, and do they have the authority to make it differently?
3. What will they now stop doing as a result?

If the honest answers are "reporting in general," "several committees," and "nothing," then you are not buying a decision improvement. You are buying speed you have no way to spend, and speed you cannot spend accumulates as cost and cynicism, not as value.

There is a further wrinkle worth naming, even from here in early 2022. The newest tools are beginning to generate — to draft code, to produce text and analysis on request — rather than merely to predict, and the temptation will be to believe that this time the productivity finally arrives. It will not, unless we learn the lesson the analytics era is still teaching us. A tool that generates answers faster is still, however impressive, a

supply-side tool. It will meet the same wall: authority, consequence, commitment. Generation without a decision to feed is simply the paradox at a higher clock speed.

6. The efficient way to stand still

Return, one last time, to the Monday room and its forty-minute pack. That number — three days down to forty minutes — is the achievement everyone will put on a slide. But it is the wrong achievement. The achievement would have been ending the failing workstream in October, when the data first said so. The tools got faster. The room did not get braver.

We are, as a profession, fluent in acceleration and clumsy at resolution. We have made the supply of answers almost free and left the cost of deciding precisely where it has always been — in the awkward human territory of authority, risk, and will. Until the decision itself changes, a faster output is not progress toward the goal. It is a more efficient, more expensive, more beautifully instrumented way of standing exactly still.