

BEYOND FASTER DRAFTS

The First Draft Was Never the Expensive Part

What organisations get wrong when AI fills the blank page — and the judgement they are quietly automating away

Giovanni Leonardi

January 2026



Contents

1. Executive Summary	3
2. The Seduction of the Filled Page	3
3. What Knowledge Work Actually Cost	4
4. Why the Cheapest Task Looks Like the Whole Job	5
5. The Fault Line Under Professional Services	6
6. The Distance Between Intent and Result	7
7. Rebuilding the Expensive Part	7
8. The Last Judgement	8

The First Draft Was Never the Expensive Part

Transformation · No. 432
January 2026

*"We automated the cheap column and told ourselves
we had transformed the expensive ones."*

1. Executive Summary

The generative tools that reached every desk over the course of 2025 do one thing with startling competence: they produce the first draft. A memo, a financial model, a client deck, a research synthesis — the blank page, for two centuries the reliable bottleneck of knowledge work, now fills itself in seconds. Most organisations have read this as the transformation. It is not.

Drafting was never the expensive part of knowledge work. Judgement was. The blank page was cheap to fear and dear to overcome only because a person's time was the currency; the genuinely costly act was always the one that came after the draft — deciding which of the plausible things on the page was actually true, and standing behind that decision when it mattered. This essay argues that the pattern now spreading through every knowledge-intensive organisation — treating "AI writes the first draft" as the whole of the change — mistakes the cheapest task in the chain for the most valuable one. The result is a widening gap between the transformation leaders believe they are buying and the one they actually get. That gap is not an implementation failure to be fixed with better prompts or another pilot. It is structural, and it is worth understanding before we automate our way past the very practice through which judgement was learned.

2. The Seduction of the Filled Page

Consider a scene that has become ordinary in the space of a single year. An analyst arrives on a Monday with a request in her inbox: a nine-page briefing on a market she has never covered, wanted by end of day. Eighteen months ago this was a full day's work and most of a second — the reading, the structuring, the wrestling of a blank document into something with a spine. This Monday she has a serviceable draft before her coffee is cool. It is well organised. It cites the right categories of evidence. It reads, on a first pass, like something a competent colleague produced.

The relief is real, and so is the trap. Because the thing she now holds is not a briefing. It is an artefact that looks exactly like the output of expertise while containing none of the expertise itself. Somewhere in those nine fluent pages is a confidently stated figure that is wrong, an inference that does not hold, a framing borrowed from an adjacent market where it does not apply. The draft has not removed her hardest work. It has disguised it —

dressed the unresolved questions in the clothing of finished answers, and handed her the job of undressing them again under the same deadline, now with the clock already half spent on generation she no longer did herself.

This is the seduction. The filled page arrives wearing the costume of completion. And because it arrives so fast, and looks so finished, the organisation around the analyst reads the speed as the whole story. The briefing went out in a morning; the transformation, surely, has happened. What the organisation cannot see — because nothing in its dashboards is built to see it — is that the expensive work did not disappear. It moved, and it may have grown.

The blank page was a visible cost, so we celebrated its removal. Judgement was an invisible cost, so we assumed it came free. The whole error of this transformation lives in that asymmetry.

3. What Knowledge Work Actually Cost

To understand why "AI writes the first draft" delivers so much less than it promises, it helps to take the thing we call knowledge work apart. Beneath the single word sit three quite different activities, and they were never equally expensive.

The first is generation — turning a brief into a first articulation. Assembling the structure, producing the prose, building the initial model, drafting the clauses. This is the part that felt hard, because it consumed visible hours and stared back at you from an empty screen. It is also, it turns out, the part a large language model does most naturally, because generation is fundamentally a problem of plausible construction, and plausible construction is exactly what these systems are built to do.

The second is discrimination — the judgement that separates the true from the merely plausible, the load-bearing claim from the decorative one, the option that survives contact with reality from the one that reads well in a deck. This is where domain expertise actually lives. It is slow, tacit, and largely invisible, because a good practitioner does it almost without narration: they simply know that a number is off, that an argument has a hole, that the obvious answer is a trap.

The third is accountability — the willingness to attach one's name and standing to a decision and to carry the consequences if it is wrong. This has never been automatable, because it is not a task at all; it is a relationship between a person and an outcome.

Activity	What it is	Who now does it	What it costs
Generation	Producing the first articulation	Increasingly the machine	Collapsing toward zero
Discrimination	Separating true from plausible	Still the human	Unchanged, arguably rising

Activity	What it is	Who now does it	What it costs
Accountability	Standing behind the decision	Only ever the human	Unchanged

Set out this way, the miscalculation becomes visible. Generative tools have driven the cost of the first column toward zero. But the value of knowledge work was never concentrated in the first column. It sat in the second and third – and those have not moved. Worse, discrimination may have become more expensive, not less. When drafts were slow and human, their errors were the familiar errors of a tired colleague, and a reviewer knew where to look. Machine drafts fail differently: fluently, confidently, and without the tells that experience taught us to read. Reviewing a plausible-but-wrong page written by a person who was thinking is one task; reviewing a plausible-but-wrong page written by a system that was not thinking at all is a harder and less forgiving one.

We automated the cheap column and told ourselves we had transformed the expensive ones.

4. Why the Cheapest Task Looks Like the Whole Job

If the error is this clear on inspection, the interesting question is not whether organisations are making it but why the mistake is so stable – why it persists across firms and sectors that are, individually, full of intelligent people. The persistence is the tell that something structural, not merely foolish, is at work. Four forces hold the pattern in place.

- The measurement gap. We can measure the time saved on generation – it is concrete, immediate, and flattering. We cannot easily measure the quality of discrimination, because judgement resists the dashboard. Faced with a benefit we can count and a cost we cannot, organisations optimise the thing they can see. Drafting hours fall by seventy per cent and the number goes on a slide; the slow erosion of decision quality shows up nowhere, because no one was ever metering it.
- The vendor and advisory narrative. An entire market has an interest in the story that the transformation is the productivity of generation. Tool providers price on seats and time saved. Advisory firms, themselves under pressure, sell programmes framed around output and efficiency because those are the terms buyers already understand. The language available to describe the change was written by the parties who profit from the narrower reading of it.
- Organisations were already built around volume. Long before these tools arrived, most knowledge work was managed as though throughput were the point – utilisation, output, turnaround. A machine that increases throughput therefore slots perfectly into the existing incentive structure and confirms it. The transformation that would actually matter – reorganising around decision quality rather than output volume – requires dismantling the very metrics that make the current transformation look like a triumph.
- The deskilling loop is invisible on a quarterly clock. The cost of thinning out judgement does not arrive this year. It arrives in five, when the cohort that never learned to draft is asked to review. Because the bill is deferred and diffuse, it never appears in the case that justifies the change. Every force above points the same way:

toward counting the saving now and discovering the cost later.

None of these is irrational at the level of the individual actor. That is precisely why the pattern is so durable. Each participant — the manager optimising a visible metric, the vendor selling a countable benefit, the firm honouring its existing incentives, the executive answering to this year's numbers — is behaving sensibly. The mistake is emergent. It belongs to the system, which is what makes it a transformation problem rather than a competence problem.

5. The Fault Line Under Professional Services

Nowhere is this fault line sharper than in the professional services that sold knowledge work as their entire product. The consulting, legal, audit, and advisory firms were built on a specific machine: the leverage pyramid. A wide base of juniors did the drafting — the research, the first models, the initial memos — under the review of a narrow apex of seniors who supplied the discrimination and carried the accountability. The economics of the firm and the development of its people ran through the same structure.

The generative tools attack the base of that pyramid directly. The work that justified a large cohort of juniors is exactly the generation that now costs almost nothing. The immediate commercial logic is obvious and is already being acted upon across the sector: fewer juniors, higher margin, faster delivery. And here the sector walks into a trap of a particularly cruel design, because the base of the pyramid was never only a production line. It was the training engine. The senior who supplies judgement in year twelve learned that judgement by drafting badly in year one, being corrected, and slowly internalising the difference between plausible and true. Remove the drafting and you have not merely cut a cost. You have unplugged the apparatus that manufactures the only thing the firm ultimately sells.

The honest counter-argument deserves its full strength, because the pessimistic reading is not automatically correct. Every previous tool that collapsed the cost of production provoked the same alarm and the same forecasts of hollowed-out capability. The spreadsheet was going to end financial literacy; the word processor was going to end the discipline of writing; the search engine was going to end the retention of knowledge. In each case the work reorganised, the apprenticeship reformed around the new tool, and the profession grew rather than withered. On this reading, the pyramid was an inefficient way to train people that we mistook for the only way, and juniors freed from mechanical drafting will simply learn judgement earlier and at a higher level, by reviewing and directing machine output instead of producing raw material by hand.

This is the strongest version of the optimistic case, and it is not empty. But it rests on an assumption that this transformation, unlike its predecessors, may not honour. The spreadsheet and the word processor automated the mechanical layer beneath judgement while leaving the judgement-forming activity intact — the analyst still decided what went in the cells and why. What is being automated now is the drafting itself: the very act through which discrimination was historically rehearsed. You learn to tell true from plausible by generating the plausible and being shown where it failed. A junior who directs a machine to produce the draft, and reviews an output already dressed as finished, is doing a genuinely useful thing — but it is not the same thing, and it is not obvious that it

builds the same faculty. Reviewing is not drafting; correcting a confident machine is not the same discipline as being the one who was confidently wrong and had to feel it. The optimistic case may prove right. But it is a bet on a substitute apprenticeship that no one has yet demonstrated, made by firms whose quarterly incentives push them to stop paying for the old one immediately.

6. The Distance Between Intent and Result

Stand back and the gap between what this transformation intends and what it delivers comes into focus. The intent, stated in a hundred board papers over the past year, is some combination of faster, leaner, and better. The result, observed across enough organisations to constitute a pattern, is faster, leaner, and quietly no better — sometimes worse — in the dimension that matters most.

Consider the composite arithmetic, drawn from what recurs rather than from any single case. A team's drafting time on a standard deliverable falls from roughly six hours to twenty minutes — a saving so large it dominates every conversation. But the review time on that same deliverable rises, because the reviewer can no longer trust the draft's provenance and must now interrogate every load-bearing claim rather than skim a colleague's reasoning. Where a senior once spent perhaps forty minutes checking work whose method they understood, they now spend ninety minutes hunting for the confident error in work whose method was no method at all. The headline shows a collapse in production time. The ledger that no one keeps shows review cost rising and, more corrosively, a slow drift in the rate at which flawed claims survive into final decisions — because under time pressure, some of that ninety-minute interrogation simply does not happen, and the fluent draft goes out very nearly as it arrived.

The intent was to move the human up the value chain — to free people from generation so they could concentrate on judgement. The reality, absent deliberate design, is that the human is pushed up the value chain into a role for which the old ladder no longer prepares them, while the organisation counts the saving at the bottom and never books the cost at the top. Faster and leaner are true. Better was assumed, and assumption is not a control.

“Faster and leaner were purchased. Better was assumed — and an assumption has never yet reviewed a document.”

7. Rebuilding the Expensive Part

If the diagnosis is that we have automated the cheap task and neglected the expensive ones, the response is not to refuse the tools — that argument was lost the moment the first draft got good, and rightly so, because the generation was genuinely worth

automating. The response is to take seriously the work the automation exposed rather than the work it removed.

Three commitments distinguish the organisations beginning to get this right from those merely getting faster.

1. Measure the expensive column, not the cheap one. If discrimination is where the value now sits, it must become visible. That means metering the things the dashboard has always ignored: the rate at which flawed claims reach decisions, the time and quality of review, the proportion of machine drafts that are materially rewritten versus waved through. What gets measured gets defended; judgement will not be protected while it remains the one thing no one counts.
1. Treat the draft as a hypothesis, not an answer. The single most important cultural shift is to strip the machine draft of its false authority – to build the habit, and the workflow, that receives a generated artefact as a confident claim to be tested rather than a near-finished product to be polished. This is a change in posture more than in tooling, and it is the opposite of the reflex the tools naturally induce, which is to trust what reads as finished.
1. Protect the apprenticeship on purpose. If the base of the pyramid was the training engine, then removing it without a deliberate replacement is a decision to stop manufacturing senior judgement – a decision most firms are making by accident, through the aggregate of sensible local cuts. The organisations that will have judgement to sell in a decade are the ones treating its formation as something to be designed and funded now: giving juniors real problems to be wrong about, having them draft before they direct, and making the review relationship a teaching one rather than a transaction.

None of this is a brake on the transformation. It is the difference between a transformation that compounds and one that hollows. The tools are not the risk. The risk is adopting them with a model of knowledge work that was always wrong about where the value lived, and letting that wrong model decide what we keep and what we let wither.

8. The Last Judgement

We have automated the first draft. We have not automated the last judgement – the act of deciding which of the plausible things on the filled page is actually true, and standing behind that decision – and it is not clear we ever can, because that act is finally a matter of a person and their accountability, not a task to be dispatched. What we have quietly done, in our enthusiasm for the filled page, is narrow the foundation on which that last judgement stands, by automating away the practice through which it was learned.

The organisations that thrive in the next several years will not be the ones that generated the most drafts the fastest. Everyone will have that; it will be a commodity by the time the current excitement cools. They will be the ones that understood, early, that the first draft was always the cheap part – and that the transformation worth having was never about producing more, but about protecting and rebuilding the expensive, invisible, irreplaceable work of knowing which of it to believe.