

BEYOND THE PLAUSIBLE DRAFT

The Compliance Acceleration — and the Assurance It Leaves Behind

Generative AI, regulatory reporting, and the widening gap between drafting a disclosure and being able to prove it.

Giovanni Leonardi

June 2024



Contents

1. Executive Summary	3
2. The Function That Was Waiting For This	3
3. Why The Fit Looks Perfect	4
4. The Paradox At The Centre	4
5. Why We Accelerate Anyway	5
6. The Distinction That Actually Matters	6
7. The Strongest Objection, Answered	7
8. The Gap Between Intent And Reality	7
9. What Compliance Actually Wanted	8

The Compliance Acceleration — and the Assurance It Leaves Behind

Transformation · No. 321

Originally written June 2024 | Re-edited and re-rendered for the WhereBeyond Edition, 2025

*“A regulator does not reward a well-written disclosure;
it accepts a defensible one.”*

1. Executive Summary

In 2024, a great many organisations pointed generative AI at their compliance and regulatory reporting functions first — before customer service, before engineering, before the commercial edges of the business where the returns look larger. The instinct is easy to understand. Regulatory reporting is text-heavy, high in volume, bound by rules, and expensive to staff. It reads, on paper, like the ideal first customer for a machine that produces fluent language at near-zero marginal cost.

This essay argues that the instinct is half-right in a way that is more dangerous than being simply wrong. Regulatory reporting is indeed saturated with language work that a capable model can accelerate. But the purpose of that work is not to produce language. It is to produce proof. A regulator does not reward a well-written disclosure; it accepts a defensible one. And the property that generative models optimise for — plausibility, the quality of reading as though it were true — is not merely different from provability. In this domain it is close to its opposite.

What follows is a reflection on why the acceleration happened anyway, what structural forces keep it moving even as the pilots quietly under-deliver in production, and where the genuine value actually sits: not in letting the model attest, but in using it to assemble the evidence a human still has to sign.

2. The Function That Was Waiting For This

Consider the shape of a regulatory reporting cycle as it is actually lived. The quarter closes. Source systems are drained into the reporting layer. Numbers are reconciled, broken, reconciled again. And around those numbers a small army assembles language: the narrative sections of a capital adequacy disclosure, the commentary that accompanies a liquidity return, the rationale in a suspicious-activity referral, the control descriptions in a model validation pack, the attestations that say, in effect, we looked, and this is right. Much of it is drudgery — high-stakes drudgery, drafted under deadline by expensive people who would rather be doing analysis.

So when generative tools arrived in force, compliance and reporting leaders did not have to be persuaded. Many of them felt they had found the use case they had been waiting years for. Here, at last, was a technology aimed squarely at the part of the job everyone privately resented: the writing. The reaction I saw most often was not caution but relief — a sense that the machine had come to take away exactly the work people were happy to lose.

That relief is the beginning of the problem, because it locates the value in the wrong place. The writing was never the point. It was the visible residue of something else.

3. Why The Fit Looks Perfect

Four properties make regulatory reporting look like the natural home for generative AI, and each of them is genuinely present.

- It is voluminous. Large institutions file hundreds of returns across dozens of jurisdictions, and the surrounding narrative runs to thousands of pages a year. Volume is where automation economics work.
- It is textual. Unlike a trading algorithm or a pricing model, a great deal of the compliance artefact is prose — explanations, justifications, policies, memoranda. Prose is what these models do.
- It is rule-bound. Reporting sits on top of detailed, published rulebooks. It is tempting to assume that where the rules are explicit, a model can simply be told to follow them.
- It is expensive. The people who do this work are scarce and costly, and the regulatory headcount line has only ever grown. Anything that promises to bend that curve gets an audience with the board.

Put those four together and the business case writes itself. A team drafting the narrative for a Pillar 3 disclosure, or the commentary threaded through a COREP return, can watch a model produce forty pages in the time it takes to fetch a coffee. The demonstration is genuinely astonishing, and the astonishment is real. It is also, almost always, measuring the wrong thing.

4. The Paradox At The Centre

Here is the tension the whole essay turns on. The value a regulator, an auditor, or a board audit committee places on a compliance artefact has almost nothing to do with how well it reads. It has everything to do with whether it can be stood behind — traced to a source, reconciled to a system, defended in a conversation that begins with the question how do you know? Compliance is, at bottom, an apparatus for the manufacture and preservation of provable statements.

Generative models manufacture something that looks identical from the outside and is different in kind on the inside. They produce the most probable continuation of a prompt — language shaped to be plausible. Most of the time, in a well-grounded system, plausible

and true coincide. The difficulty is that compliance is one of the few domains where the entire value lives in the gap between them.

A model that is right ninety-four per cent of the time is a triumph in most settings and a hazard in this one, because the purpose of a regulatory return is precisely to be accountable for the six per cent. The exception is not noise around the signal. In compliance, the exception is the signal.

I have watched this play out in a way that is worth describing concretely, because the abstraction hides the mechanism. A reporting team adopts a model to draft the narrative around a capital return. It is a success by every measure taken in the first week: drafting time for the commentary falls by perhaps seventy per cent. Then a reviewer, doing what reviewers do, notices that a figure quoted in the prose does not tie to the source ledger. It is close — plausibly close — but wrong. And now something expensive happens. Because the reviewer can no longer trust any number the model transcribed, every figure in all forty pages must be re-verified by hand. Trust in an assurance artefact is binary; a single fabricated reconciliation poisons the whole document. Drafting time went down seventy per cent; review time went up more than that; net cycle time was flat at best, and the confidence with which anyone signed the return was lower than before the model was introduced.

That is the pattern beneath the enthusiasm. The technology reliably compresses the part of the work that was never the constraint — the drafting — and quietly inflates the part that always was: the assurance.

5. Why We Accelerate Anyway

If the mismatch is this legible, why does the acceleration continue? Because the forces sustaining it have little to do with whether it works and a great deal to do with how organisations are wired.

The first force is board pressure. In 2024 no board wants to be the one without an AI story, and compliance — large, cost-heavy, and full of obvious language work — is the easiest place to point at and say we are modernising. The narrative is irresistible precisely because it is legible to non-specialists.

The second is vendor incentive. Every reporting platform, every governance-risk-and-compliance suite, every document workflow now ships a copilot, and each is sold on the same demonstration: watch it draft. The demonstration is honest about speed and silent about assurance, because assurance does not demo well.

The third, and the most corrosive, is a measurement gap. We instrument the things that are easy to count — drafts produced, hours notionally saved, tickets closed — and we do not instrument the thing that matters, which is assurance: defects caught before filing, attestations that survive challenge, findings avoided. When the metric is activity, a tool that produces more activity always looks like progress, even when it is manufacturing rework downstream.

Pilots are judged on speed and spectacle; production is judged by the regulator and the person who signs. The two are graded on opposite criteria, which is why the pilot dazzles and the rollout disappoints — and why the disappointment is so often misread as an implementation problem rather than a category error.

There is a particular irony worth naming. The very function whose job is to demand that other parts of the business explain their models — to insist on validation, on documented lineage, on the discipline that model-risk guidance has required since well before this wave — is now under pressure to deploy into itself an opaque model it cannot fully explain. Compliance is being asked to hold itself to a lower standard than it holds everyone else. With the AI Act agreed in Brussels this spring and now awaiting its staged entry into force, and with operational-resilience obligations tightening across the sector, that inconsistency is not going to age well.

6. The Distinction That Actually Matters

The way through is not to reject the technology; it is to draw one hard line and hold it. The line runs between AI as an assistant and AI as a system of record.

An assistant drafts, summarises, triages, proposes, and does a competent first pass — all subordinate to a human who remains the author of record. A system of record is the authoritative source of a fact. The single most important rule in applying generative AI to regulatory reporting is that the model may be the former and must never be the latter. It may write the words around a number; it must never be the origin of the number. Figures come from deterministic systems with lineage you can walk backwards; the model arranges language around them, and a human attests.

Where the model earns its place	Where the model must not go
Drafting narrative around figures sourced elsewhere	Producing or "estimating" the figures themselves
Summarising long rulebooks for a human to verify	Being the final authority on what a rule requires
First-pass triage of alerts and exceptions	Closing or dispositioning cases without human sign-off
Reconstructing and proposing evidence trails	Serving as the evidence trail of record
Explaining a control to a human reviewer	Attesting that the control operated

This is the discipline sometimes described as building the controls around the model rather than trusting controls inside it. You do not ask the model to be reliable; you architect a system in which its unreliability is contained. Prompts and outputs are logged so the process is itself auditable. Every generated figure is reconciled against a deterministic source before a human ever sees it. And critically, you never let the same model both produce a document and check it — that is the compliance equivalent of

keeping a second set of books, marking your own homework with a copy of your own hand.

7. The Strongest Objection, Answered

The serious counter-argument deserves its strongest form, not a caricature. It runs like this: You are describing a 2024 limitation, not a permanent condition. Models are improving fast; retrieval grounding already tethers them to real sources; tooling will keep closing the gap between plausible and true. And regulators are pragmatic — they will adapt their expectations as the technology matures. You are mistaking the awkward early middle of an adoption curve for a law of nature.

Part of this is right, and it is worth conceding cleanly. The drafting, summarising, and triage value is real, it is not a mirage, and it will grow. Grounding genuinely reduces fabrication. Anyone who dismisses the whole enterprise is going to look foolish in a few years.

But the objection answers a question compliance is not asking. Regulation does not ask is this true? It asks can you demonstrate why you believed it, and who is answerable if it is wrong? Those are questions about attributable reasoning and human accountability, and they are not dissolved by accuracy. Imagine the limit case: a perfectly accurate oracle that returns the correct answer every time but cannot show its reasoning and cannot be held to account. That oracle is still inadmissible as a system of record, because an unexaminable correct answer fails the actual test. The bottleneck was never plausibility versus truth in the first instance; it was provability and answerability, and those do not improve on the same curve as model accuracy. Better models make a better assistant. They do not make an accountable author, because accountability is a property of persons and institutions, not of predictions. The mistake in the optimistic case is one of category, not of degree.

8. The Gap Between Intent And Reality

This is where transformation intent and transformation reality part company. The intent, stated in every business case, is efficiency: faster reporting, fewer hands, lower cost. The reality that arrives is subtler and, understood properly, more valuable.

What actually happens in the organisations that persist past the disappointing pilot is not that the reporting team shrinks. It is that the work moves up. The hours that used to go into drafting move into judgement — into interrogating the exceptions, testing the controls, defending the numbers. The analyst who spent Thursdays writing narrative now spends them deciding which of two hundred model-surfaced anomalies actually matter. That is a real transformation. It is simply not the one the business case described, and because it does not reduce headcount in the way the slide promised, it is often experienced as a failure even when it is a success.

The remedy is to change what we measure. As long as the metric is hours saved on drafting, the technology will look like it is winning while assurance quietly erodes. The

metric that tells the truth is closer to assurance per hour — how much defensible confidence the function produces per unit of effort. Reframed that way, the model's contribution becomes honestly visible: it is enormous in the connective and investigative work, and roughly zero, or negative, in the act of attestation itself.

9. What Compliance Actually Wanted

Step back far enough and the whole episode resolves into a single misunderstanding about what was being asked for.

Compliance never wanted a faster writer. It wanted a more trustworthy memory — a way to know, quickly and completely, what happened, where the number came from, which control operated, and whether the story it is about to tell a regulator will hold. The drafting was only ever the surface expression of that deeper need, and by automating the surface first we mistook the residue for the substance.

The genuine prize sits exactly where the enthusiasm has not been pointing: in evidence assembly, in reconciliation at scale, in testing controls across a whole population rather than a sampled handful, in reconstructing lineage that used to take a week, in triaging the flood of alerts down to the few that a human must judge. These are retrieval, connection, and checking problems — the machine is superb at them — and none of them require the model to assert a single thing on its own authority.

*“Automate the search for the truth. Do not automate
the assertion of it.”*

That is the line the compliance acceleration keeps crossing, and it is the line worth holding. Point the machine at finding the evidence, assembling it, stress-testing it, and surfacing what does not fit — and let a human, who can be asked how do you know and can answer, remain the one who signs. Do that, and the technology stops being a faster way to produce plausible documents nobody can quite trust, and becomes something the function has genuinely always wanted: a way to be more certain, faster, of things it can actually prove.