

BEYOND THE FASTER DRAFT

The Compliance Acceleration: Why AI Sped Up the Wrong Half of Regulatory Reporting

What a year of pointing generative models at regulatory returns taught us about where the real bottleneck actually lives

Giovanni Leonardi

May 2024



The Compliance Acceleration: Why AI Sped Up the Wrong Half of Regulatory Reporting

Transformation · No. 380

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

"You cannot attest to a black box."

1. The Demo and the Draft

Last autumn I sat in a room while a vendor fed a general-purpose language model the previous quarter's regulatory return and asked it to write this one. In under a minute it produced the qualitative narrative — the commentary that sits above the capital numbers, the movements explained, the tone measured and fluent, the house style more or less intact. The room was senior and, on the way in, sceptical. It applauded. Someone said the quiet part out loud: that used to take a fortnight.

It did not, though. The drafting took a fortnight because everything upstream of the drafting took a fortnight. The paragraph a model can now generate in seconds was never the expensive part of the return. It was the last part — the visible tip of a great deal of invisible reconciliation, checking, challenge and sign-off — and it was expensive only because everything beneath it had to be true before anyone dared write it down.

That gap — between what the demonstration accelerated and what the work actually costs — is the most useful thing I have learned watching organisations reach for generative AI in regulatory reporting over the past eighteen months. We have been told, insistently, that this technology will transform compliance. In one narrow sense it already has. In the sense that matters, most of what I have seen accelerates the wrong half of the pipeline, and a few firms have quietly made themselves harder to defend in the bargain.

The argument of this piece is simple to state and, apparently, hard to act on: the bottleneck in regulatory reporting was never the speed of writing. It was the defensibility of what gets written — the ability, when a regulator calls, to stand behind every figure and name who checked it. Point AI at the prose and you automate the cheap, safe, visible five per cent. The acceleration worth having sits at the other end entirely.

2. What a Return Actually Costs

It is worth being concrete, because the whole misunderstanding lives in an unexamined sense of where the effort goes. Take a substantial quarterly return — a capital or liquidity submission, a set of prudential disclosures, the qualitative and quantitative package a mid-sized regulated firm files as a matter of course. Break its production into labour and the shape is consistent across the places I have seen it done well and badly alike.

Activity	Roughly, per quarterly cycle	What it actually involves
Sourcing and reconciling data	~20 person-days	Pulling figures from ledgers, risk systems and sub-ledgers that disagree, and making them agree
Control checks and validation	~8 person-days	Running the reconciliations, chasing the breaks, evidencing that each control fired
Internal review and challenge	~6 person-days	Second line reading the numbers, asking why a ratio moved, sending it back
Sign-off and attestation	~4 person-days	The accountable person satisfying themselves enough to put their name to it
Drafting the narrative	~2 person-days	Writing the words that explain the numbers

The figures are illustrative, not audited, but the proportions are not controversial to anyone who has lived inside the cycle. Roughly forty person-days of work, of which the writing — the thing the demo compressed to a minute — is around one part in twenty. Automate it perfectly and you have removed five per cent of the effort: the one part that carried almost none of the risk and almost none of the delay.

Worse, you have removed it at the safest point in the chain and left the reviewer exactly where they were. The narrative still has to be checked against the numbers. The numbers still have to be traced to their sources. A model that writes the commentary faster does not shorten the reconciliation, does not close the breaks, and does not relieve the second line of a single hour of challenge. It just means the words arrive sooner and wait longer for everything else to catch up.

3. The Bottleneck Is Accountability, Not Fluency

There is a deeper reason the drafting was never the constraint, and it is not about effort at all. Regulatory reporting is not a writing task with a compliance flavour. It is an accountability instrument. Somewhere at the end of the chain a named individual — under the Senior Managers regime in the United Kingdom, and its equivalents elsewhere — has to attest that the submission is complete and correct to the best of their knowledge. That attestation is the product. Everything else is scaffolding for it.

You cannot attest to a black box. When the accountable person signs, they are not vouching for the elegance of the prose; they are asserting that they can answer the two questions a regulator actually asks — where did this number come from and who checked it — for any figure on the page. A model that produces a fluent narrative gives them nothing to attest to. If anything it takes something away, because it introduces a confident, well-formed artefact whose provenance the signer did not witness.

A rough draft that is visibly unfinished invites scrutiny. A polished draft that is subtly wrong evades it. Fluency is not a neutral gift in a control environment — it is a way of passing the eye that the eye was relying on.

This is the failure mode I saw most often and worried about most. The models are good at register. They produce commentary that reads exactly like commentary a competent analyst would write, which means a tired reviewer at the end of a reporting week is precisely the person least equipped to catch the one figure that has been transposed, or the movement that has been explained plausibly and incorrectly. The technology is strongest at exactly the thing — surface plausibility — that our controls were never designed to challenge, because until now plausible prose was a reasonable proxy for careful work. That proxy has quietly broken, and most control frameworks have not noticed.

4. Where the Acceleration Actually Lives

None of this is an argument that the technology has no place in the reporting function. It is an argument about which end of the pipe. The prize — the genuine acceleration — is not producing the narrative faster. It is compressing the distance between a regulator's question and a defensible answer.

Consider what that distance is made of. A supervisor asks why a particular exposure moved between two submissions. Answering means locating the figure, tracing it back through the reconciliation to the source systems, retrieving the control evidence that it was checked, and reconstructing the commentary that explained it at the time. In most firms this is a scramble — days of people who know where the bodies are buried reconstructing a lineage that was never assembled in one place. The cost is not in the answering; it is in the archaeology.

That archaeology is exactly the kind of retrieval, correlation and assembly the current models are genuinely good at — and where their errors are cheap, because a human is checking the result against ground truth they can see. Pointed there, the technology earns its place:

- Lineage on demand — assembling, for any figure, the chain from submitted number back to source, so the answer to where did this come from is minutes away rather than days.
- Break triage — reading reconciliation exceptions and clustering them by likely cause, so the eight days of control work start with a ranked list rather than a blank one.
- Evidence assembly — gathering the control artefacts that prove each check fired into the pack the accountable person actually needs to sign with confidence.
- Consistency challenge — flagging where this quarter's narrative and this quarter's numbers disagree, which is the reviewer's job and the one place a model's tirelessness is pure gain.

Notice what these have in common. In every case the model is pointed at evidence, not assertion; its output is checked against something verifiable; and it shortens the accountable person's path to a defensible position rather than handing them a fluent

artefact they must now defend. That is the acceleration worth building. It is also, tellingly, the one no vendor demo opens with, because it does not produce an applause moment. It produces a shorter Friday.

5. The Case for Pointing It at the Prose Anyway

I should put the strongest version of the opposing view, because it is not a foolish one and I have heard capable people make it.

The case runs like this. Drafting toil is real. Analysts do spend hours turning a spreadsheet into readable commentary, quarter after quarter, and much of it is mechanical misery that burns out good people and adds no judgement. A copilot that produces a competent first draft lets the analyst start from eighty per cent and spend their scarce attention on the twenty per cent that needs a human. Multiply that across every return a large institution files and the saved hours are not trivial. And regulators themselves increasingly expect firms to be engaging seriously with these tools; standing aside can look less like prudence than paralysis. Start where the friction is visible, prove the value, and extend from there.

I do not dismiss this. The copilot genuinely helps the individual analyst, and easing that grind is a real good. But three things have to be said against it.

1. It relieves toil, not risk, and it is risk that gates the cycle. The saved drafting hours are real but they sit off the critical path; the return does not ship sooner because the words arrived sooner. You have made the cheap step cheaper and left the expensive one untouched.
2. The saved minutes are easily eaten by the review they create. A draft the reviewer did not watch being built is a draft the reviewer must now reverse-engineer. In more than one instance I have seen review time lengthen after a copilot was introduced, because the second line — rightly — stopped trusting a provenance it could no longer see.
3. Worst, and quietest: it relocates accountability without anyone deciding to. The analyst trusts the model, the reviewer trusts the analyst, the signer trusts the reviewer, and the number at the bottom of the chain is now one that no human actually generated or fully traced. Each link is individually reasonable. The chain has a hole in it.

“The productivity was real and the risk grew at the same time, which is the most dangerous combination a control function can face — because the felt experience is entirely of progress.”

That third point is the one that keeps me up. A control environment degrades most dangerously when it degrades invisibly, and a tool that makes everyone feel faster while diffusing responsibility for the numbers is a near-perfect instrument for invisible

degradation. The firms that started at the prose were not wrong to see the toil. They were wrong about what the toil was costing them, and the fluency of the result made the miscalculation hard to see.

6. What the Firms Who Got It Right Did Differently

A handful of organisations I watched navigated this well, and their behaviour was consistent enough to describe as a pattern rather than luck.

They pointed the technology at evidence before they pointed it at expression. The first production use was lineage retrieval or break triage, not narrative generation — the unglamorous end, where a wrong answer is caught by the human who asked the question. They treated the model as a model, which is to say they brought it inside the model risk discipline they already had. The validation, monitoring and documentation expectations that have governed quantitative models for well over a decade are not a bad fit for a language model in a regulated process; they are close to exactly the right fit, and the firms who reached for that existing rulebook were spared a great deal of improvisation. They kept the human genuinely in the loop at the point of accountability — not as a rubber stamp on an artefact they could not inspect, but as the party who assembled the defensible answer with the model's help.

And they read the direction of regulation correctly. The AI legislation now moving through its final approval in Brussels, whatever one thinks of its detail, encodes an instinct that the careful firms already shared: that where these systems touch high-stakes decisions, what matters is documentation, human oversight and the ability to explain. A firm that had pointed AI at evidence and kept its accountability chain intact was already, without trying, most of the way to defensible. A firm that had automated the narrative and thinned the chain had built something it would shortly have to unwind.

7. The Acceleration Was Real. It Was Just Somewhere Else.

The compliance acceleration of the past year and a half was not a fiction. Something genuinely got faster. But the thing the demonstrations sped up — the writing — was the part that was never slow in any way that mattered, and pointing our best new tool at it felt like progress while quietly moving accountability to where no one was standing.

The work that actually gates a regulatory return is the assembly of a defensible position: the lineage, the evidence, the checked and challenged number that a named person can stand behind. That is where the days are, that is where the risk is, and — for anyone willing to skip the applause moment — that is where the acceleration was available all along. We were shown a faster draft. What we needed was a shorter distance to an answer we could defend, and the two are not the same thing at all.