

The compensating structure

Why most institutional AI work is tooling, and what the alternative actually looks like.

Naga Lalitha Kumar Mudiganti

ORCID 0009-0000-5907-3704

Draft for discussion

Ask a room of senior bankers whether AI will change their business and every hand goes up. Ask what it is being used for today and you will hear: summarising documents, drafting client correspondence, accelerating code review, answering policy questions faster. Both answers are honest. They describe entirely different things, and the distance between them is where the next decade of institutional advantage sits.

The second list is tooling. Tooling makes an existing process cheaper, faster or less unpleasant. It is worth doing — the returns are real and they compound — but it leaves the shape of the business exactly as it found it. A bank that deploys tooling perfectly across every function ends up as the same bank, operating at lower cost.

There is a test that separates the two, and it takes one question:

| What does this let us stop doing?

Tooling never answers it. Tooling makes a step faster; the step remains, and so does everything downstream of it. Transformation answers immediately, and the answer is usually something large, expensive and long-standing — something the organisation has a department for.

Why that answer comes so rarely is worth examining, because the reason is not a failure of imagination. It is that the processes most worth removing were never arbitrary. They were built as compensations.

Every large process is compensating for something unreadable

Somewhere in the history of any substantial institutional process there is a piece of information that could not be read. Not information that was missing — information that existed, that the institution held, that was often contractually precise and legally binding, but that could not be computed on at the speed or the scale the business needed.

Around that illegibility the institution built a structure. The structure is not a mistake. It is usually an elegant answer to a genuine constraint, designed by capable people who understood the problem better than any critic did afterwards. But it carries a cost, that cost is paid continuously, and — this is the part that matters — the structure long outlives the constraint that produced it.

Call it the compensating structure. It has a recognisable shape wherever it appears: a body of process, infrastructure and convention whose only real justification is that a particular thing could not be read. Nobody defends it in those terms, because by the time anyone asks, its origin has been forgotten and it has acquired owners, budgets, regulatory standing and a place in the org chart.

The genuine contribution of AI — the part that is not tooling — is that it dissolves the constraint. Not by computing faster, but by making previously illegible material computable at all. And the moment the constraint dissolves, the compensating structure is revealed for what it now is: a large, expensive apparatus maintained in order to work around a problem that no longer exists.

That is the whole thesis. What follows is three cases, from three unrelated businesses, that have the identical shape.

Three domains, one structure

I Market risk: the second machine

A bank's risk position is one thing economically. It is represented as several. Stress testing, the trading-book capital regime and internal limit frameworks each specify their own representation of overlapping quantities, on different cadences, computed by different infrastructure — because each was specified separately, and no representation was ever legible to the others.

So the bank builds a second machine. Then a third. Parallel estates computing quantities that describe the same underlying exposures, plus a reconciliation function whose entire purpose is to explain why two answers to the same question differ. The capital position that results reflects the most conservative machine rather than the true state, because nobody can demonstrate that the machines agree.

None of this was irrational. It was the only available answer when the representations genuinely could not be reconciled at scale. Make them legible to one another and a single continuously computed risk state can serve every regime, each view derived rather than separately produced.

UNREADABLE

Each regime's risk representation, to every other regime

COMPENSATION

Parallel risk estates, plus permanent reconciliation between them

STOP DOING

Building the second machine — and reconciling it

Transaction banking: the fungible dollar

Every payment carries context. Where it originated, what it was for, which counterparty relationship sits behind it, what path it travelled, what it implies about the flows that follow. Almost none of that context survives into how the flow is actually treated. A dollar arriving is treated as identical to any other dollar, because for operational purposes it has to be.

The compensation is fungibility itself: manage the aggregate, size liquidity buffers for the average, and accept the cost of holding against a distribution you cannot see. The consequence is quietly expensive. A bank holding the primary relationship at both ends of a flow — seeing its origin and its destination — extracts no more value from that position than a bank holding neither, because the advantage exists in information it cannot read.

Make the flow legible and the premise inverts. Not all dollars are equal; provenance becomes a computable property; risk is measured on the distribution rather than the mean. The relationship advantage stops being purely a relationship and becomes a data position.

UNREADABLE

The provenance, purpose and forward implication of each individual flow

COMPENSATION

Treating flows as fungible; buffers sized for ignorance

STOP DOING

Pricing a relationship advantage as though it were only a relationship

iii Private credit: the calendar

Every credit agreement defines a precise set of tests — leverage, coverage, capitalisation caps, waiver and amendment conditions. Borrowers report against them, and a chief financial officer signs the report. The information is contractual, attested, and already in the lender's possession. It is also unstructured prose, which means it cannot be computed at portfolio scale, and it says nothing at all about the eleven weeks between one report and the next.

The compensation is the calendar. Quarterly reporting, hold-to-maturity accounting, smoothed valuations, appraisal-based marks. Pricing runs on a calendar because verification runs on a calendar. The cost shows up as a benchmark whose reported volatility understates the underlying by more than an order of magnitude, collateral haircuts that jump discontinuously at reporting boundaries, and a secondary market that cannot properly form because no buyer can verify what they are buying between dates.

Read the agreements and the reports, and observability separates from the reporting cycle. The tests were always there. Only the reading was missing.

UNREADABLE

Contractual predicate state, at portfolio scale and between report dates

COMPENSATION

The quarterly cycle; smoothed, appraisal-based valuation

STOP DOING

Treating a quarterly mark as though it were a price

DOMAIN	WHAT COULD NOT BE READ	WHAT WAS BUILT INSTEAD	WHAT BECOMES UNNECESSARY
Market risk	Each regime's risk representation, to the others	Parallel risk estates and permanent reconciliation	The second machine
Transaction banking	Provenance and purpose of each flow	Fungible treatment; buffers sized for the average	Holding against an unseen distribution
Private credit	Contractual predicate state between report dates	The quarterly cycle and smoothed valuation	Pricing on the calendar

Three businesses with nothing in common — different clients, different regulators, different infrastructure, different people. The same structure in each. In every case the information already existed inside the institution. In every case what was built was a workaround for the inability to read it. And in every case the workaround is now the largest single line of cost in that part of the business.

Why this is invisible from inside a function

Each of these looks, from within its own function, like a domain problem with a domain answer. The market risk version is a regulatory architecture question. The treasury version is a liquidity management question. The private credit version is a valuation and disclosure question. Every one of them has a well-developed literature, and none of those literatures references the others.

The pattern only becomes visible after the third one, and that is not an argument for cleverness — it is an argument about vantage point. Someone who has spent a career inside one function will diagnose the compensating structure in that function as a peculiarity of that function, which is exactly how it presents.

The practical consequence is what makes it worth naming. If the structure is general, then the capability that dissolves it is transferable — the same engineering, the same governance argument, the same sequencing problem, applied to a different corpus. That is a very different investment case from three unrelated projects.

What follows for how you build

Recognising the pattern is the easy half. The three cases converge just as strongly on *how* the work has to be done, and these constraints are not stylistic — they are what determines whether anything survives contact with a risk function.

1 **Separate interpretation from calculation**

AI reads unstructured evidence and produces structured state. The financial calculation that consumes it stays deterministic, transparent and governed, with no model in the path. This is not a presentational choice. It is the difference between a proposal that passes model validation and one that does not.

2 **Provenance is the precondition, not a feature**

Every derived value carries the source it came from — the clause, the line, the document, the date. Without lineage, review is impossible, validation is impossible, and nothing built on top can ever cross an institutional boundary.

3 **Uncertainty must survive to the output**

A system that cannot return *unknown* is manufacturing precision it does not have. Surfaced uncertainty is reviewable; swallowed uncertainty is unbooked model risk, and it is the first thing a good validator will look for.

4 **Decision support long before decision authority**

Most of the value arrives while the system is still only informing people. Authority over a live decision should be sought one class at a time, lowest consequence first, and only once validation evidence has accumulated on historical data.

5 **Every stage independently worth doing**

A programme that only pays off if all of it is funded will never be started, because no single person can approve all of it. Each stage must deliver something the business would want even if the next stage is refused — which also means each stage must have a stated condition under which the work stops.

What this does not claim

- **That the compensating structures were errors.** They were correct answers to real constraints, and in most cases they are still the only thing holding the process together. The claim is about the constraint expiring, not about the judgment of the people who worked around it.
- **That any of this reduces underlying risk.** It changes when and how existing risk becomes observable. In the private credit case that distinction is a formal result rather than a caveat: the risk is identical across information regimes, and only its visibility differs. The opposite claim gets made constantly and it is false.
- **That it moves quickly.** Legibility is cheap to demonstrate and slow to institutionalise, precisely because the compensating structure has owners, budgets and regulatory standing. The technical constraint is rarely the binding one.
- **That the reading is reliable enough yet in every domain.** That is an empirical question per corpus, it is measurable, and it should be measured before anything is built on top of it rather than asserted.

Where this goes

The private credit case is the furthest developed. The underlying architecture — continuous risk pricing of confidential-borrower assets through attested factor reporting — is under peer review, and a staged implementation programme exists that begins with nothing more than reading credit agreements the institution already holds, with a measurable accuracy gate at ninety days. The market risk and transaction banking cases are specified at architecture level.

What none of the three needed, at the point where they became interesting, was new technology. Proof systems, language models and the quantitative machinery were all available. What was needed was the observation that the process everyone was trying to optimise existed only to work around something that had recently become readable.

The constraint was never capability.
It was legibility.

The institutions that treat AI as tooling will spend the next decade making their compensating structures more efficient. That is not nothing. But it is a strategy that ends with the same business, run at lower cost, still paying for a workaround whose reason has expired — and competing with someone who stopped paying for it.

Status. A point of view offered for discussion, not a proposal and not a recommendation. The three cases summarise work at different stages of development; the private credit architecture is under peer review at a journal and its quantitative results are laboratory findings under calibrated simulation rather than observed market outcomes. Domains are described generically; nothing here characterises the position, systems or performance of any particular institution.

Correspondence and the underlying working papers: ORCID 0009-0000-5907-3704.