AI Governance Letter · Issue #5

The Survivors' Number

One portfolio, one bank, one system — and the figure is always taken over the ones still there.

November 2026 · Reto Gruenenfelder, IT AI Advisory

Impressum

IT AI Advisory Sdn. Bhd. (202601023072)
B-3A-8, Icon City, Jalan SS8/39, 47300 Petaling Jaya, Selangor, Malaysia
Responsible for content: Reto Gruenenfelder
reto@itaiadvisory.com · itaiadvisory.com

Published for informational purposes only. Not legal, regulatory, or compliance advice. Full disclaimer: itaiadvisory.com/disclaimer

Dear Reader,

Almost nothing in this letter is about artificial intelligence. The mathematics was settled long before anyone trained a model, and a reader from 1960 would recognise every step. What artificial intelligence changes is not the law but how often it applies — to more of the firm, more times a day, through systems that act without anyone deciding to look. The exponent moves. The arithmetic does not.

That is worth saying at the start, because what follows looks like a technical point and is in fact a reporting one.

The Claim

A firm and its client are not looking at the same number. The figure that describes a book of portfolios and the figure that describes one portfolio are different quantities, produced by different operations, and only one of them concerns the client. Nothing on the report says which is which.

The same holds one level up, between a supervisor and a regulated institution, for the same reason and with the same consequence.

What follows is a demonstration rather than an argument.

One Game, Two Summaries

One hundred dollars. Each round, a fair coin. Heads multiplies the holding by 1.5. Tails multiplies it by 0.6. Percentages of the holding at that moment, never fixed amounts.

Two rounds produce four outcomes, each equally likely.

outcome	holding
heads, heads	100 × 1.5 × 1.5 = $225.00
heads, tails	100 × 1.5 × 0.6 = $90.00
tails, heads	100 × 0.6 × 1.5 = $90.00
tails, tails	100 × 0.6 × 0.6 = $36.00

The probability-weighted mean is $110.25, which is one hundred dollars compounded at 5% a round, twice. The typical outcome is $90.00; two of the four rows end there, a 10% loss.

Both numbers come from the same four rows. Nothing was added in cents anywhere — every holding compounded percentages. They are two summaries of one table, which is why they can be compared, and why the difference matters.

A Hundred Rounds

Over one round the gap is a curiosity. Over a hundred it is the whole story.

Roughly fifty heads and fifty tails. Multiplication does not care about order, so each heads-and-tails pair multiplies the holding by 1.5 × 0.6 = 0.9. A tenth is lost every two rounds, however they are arranged.

$$100 \times 0.9^{50} = \$0.52 \qquad 100 \times 1.05^{100} = \$13{,}150$$

> The 50% gain never recovers the 40% loss, because the loss is taken from a larger base.

Those two figures describe the same hundred rounds of the same game. One is the arithmetic mean of the multipliers, (1.5 + 0.6) ÷ 2 = 1.05, which is above one. The other is the geometric mean, √(1.5 × 0.6) = 0.9487, which is below one. A process can have a positive average and a negative trajectory at the same time, and nothing is wrong with either number.

For readers who want the derivation: the holding after T rounds is a product, not a sum. Taking logarithms turns the product into a sum, to which the law of large numbers applies. The average of the logarithms is ½(0.405465) + ½(−0.510826) = −0.052680, and exp(−0.052680) = 0.948683. Averaging the logarithm and exponentiating is taking the geometric mean. That is the whole difference in one line.

Addition Across Holders, Multiplication Across Time

The instinct that something is being added is correct. What is added is holdings, not rounds. Pooling holdings and dividing by their number is addition, and it is the right operation for that question. Compounding one holding through rounds is multiplication, and it is the right operation for that one. Both appear in the same game, each in its own direction.

> The error is not the use of an arithmetic mean. It is the use of the operation that runs across holders to answer a question about one holder through time.

Where Each Summary Belongs

A firm with 4,000 portfolios averages across them, and for the firm that is the correct instrument. The book is a population, and the mean describes it. A firm that reported anything else would be reporting badly.

A client holds one portfolio and lives along it. The client does not experience the average of 4,000 portfolios, any more than a bank experiences 2% of an enforcement action. A bank experiences one or none — and a bank that experiences one does not rejoin the other 98% next year. It is in remediation for three years.

The Relation Repeats

For the firm, a portfolio is one unit among 4,000. For the supervisor, the firm is one unit among a few hundred. The relation is identical at both levels, and at every level between

them: whoever sits above holds a population and averages across it; whoever sits below holds one object and lives along it.

There is therefore no class of participant that is a population and another that is a trajectory. Each is both — a population to the level beneath, a single path to the level above. A chief risk officer reads one institution as a path and reads its own desks as a population. A head of desk does the same again, one level down.

This is what makes the problem systemic rather than local. At every hand-over a figure computed as an average across units is passed to a recipient who will read it as a description of one unit. The error does not happen once, at the top. It happens at each step, and the figure arrives several levels away from the trajectory it is taken to describe.

Where the Repetition Stops

The sequence does not continue indefinitely. It ends at the level where the participants are joined to one another.

Run the game with players who each keep their own holding, and the typical player ends at $0.52. Ninety-nine in a hundred end below the average of $13,150, and the remaining one in a hundred carries almost all of it. Run the same game with the same coin, but pool everything across many independent players and divide it equally each round, and every player ends at about $13,150.

arrangement	outcome
each keeps their own holding	typical $0.52; nobody holds the mean
pooled and divided each round	about $13,150 each

Same game, same coin. The average becomes a quantity someone actually experiences only when the participants are joined together.

That is what the safety net does, and it is why a supervisor’s figure is legitimate about the system. Deposit insurance, resolution funds and a lender of last resort are the joining. A single institution is not joined to them in that sense: it carries its own path and its own absorbing state. The pooling is also partial — depositors are protected and shareholders are not, resolution funds are finite, and the lender of last resort is discretionary — so the realised outcome sits between the two rows above rather than at either.

Rows and Columns

Picture a grid. The rows are institutions; the columns are years. Each row is one institution’s path through time.

The supervisor reads down a column: all institutions, one year. The chief risk officer reads along a row: one institution, all years. The assumption that averaging a column gives the same answer as averaging a row has a name, and it is the only piece of vocabulary this letter will ask for. If the two agree the process is called ergodic. The processes that decide whether an institution survives are not.

Put that way it is plainly a substantive assumption rather than a technicality, and plainly one that somebody should have checked. For a column to say anything about a row, two things must hold: the statistics must be the same in 2020 and in 2030, and the rows must be interchangeable, so that your institution is a draw from the same distribution as the others. In banking neither holds. Technology and regulation move, and institutions differ structurally in exactly the ways that matter.

The Survivors' Number

The grid also shows why the two averages come apart, and it is not subtle. A failed institution's row simply stops. There is no data after failure. So every later column is averaged over the rows still running.

> The figure is not wrong. It is a true number about the institutions that are still there.

Which is a different thing from a true number about the industry, and a very different thing from a true number about you. The cross-section is systematically more reassuring than any single institution's position warrants, and it becomes more reassuring precisely as conditions worsen, because that is when rows stop.

This is not an exotic defect. It is already present in the ratios on every performance page in the building. A standard deviation and a Sharpe ratio are ensemble statistics: they describe the distribution of outcomes a strategy would produce across many runs. They are computed, in practice, on one run — and they have no concept of ending. A strategy that is destroyed does not score minus infinity. It stops being measured.

So the ratios are not being used wrongly in the sense of bad arithmetic. They are answering a question about a population, in a document read by someone who holds one path. The same sentence describes the supervisor's statistic exactly.

There is a special case of this that one corner of finance already knows and has named. Two retirees with identical average returns, in a different order, and one of them runs out of money. It is called sequence-of-returns risk, it has its own spreadsheet, and it is treated as a retirement-planning quirk rather than as one instance of a general law about anything that compounds.

What the Cross-Section Can Measure

Three quantities are routinely collapsed into one number, and holding them apart is most of the work.

quantity	what it is	in the game
the expectation	what an infinite population would average to	$13,150
the cross-section	what a real supervisor observes	far below it — one sample of twenty averages $8.88
the trajectory	what one institution lives through	$0.52

The distribution behind those figures is exact, not approximated. A path with k heads ends at $100 \times 1.5^k \times 0.6^{100-k}$, and k is binomial. The mean sits at $13,150 and the median at $0.52, because the mean lives in a handful of rare, lucky paths: 99% of paths end below it, and the top 1% carry 98% of it. The coefficient of variation is about 4,590.

A sample mean within 50% of the truth needs roughly the square of that, divided by the square of the tolerance — about 84,000,000 institutions. Within 10%, about two billion. No supervisor on earth observes a population of that size, and that is before anything is correlated.

Correlation makes it worse in a way that is worth stating precisely, because it is counter-intuitive. Suppose each institution follows one common outcome with probability ρ and otherwise draws its own — shared foundation models and shared data vendors, in other words. The marginal probability for any single institution is unchanged, so ρ does not touch its own fate at all. What ρ changes is how institutions relate to one another: two of them coincide with probability ρ^2, which is the pairwise correlation, and the effective number of independent observations tends to its reciprocal.

shared-model fraction	300 institutions	10,000	one billion
none	300	10,000	1,000,000,000
70%	2.0	2.0	2.0

With 70% of the market on shared models, a single year's cross-section of growth rates carries the information of about two independent institutions, however many are supervised. For the positions themselves the correlation fades over a long horizon, and the dispersion above becomes the binding limit; the paper sets out both. Frequency is not corroboration — the argument of the September letter, reappearing one level up.

A caution that must travel with these figures: this is a toy process, not a model of banking. The numbers illustrate the mechanism, not the industry. What is robust is the direction — small populations and correlated paths make cross-sectional averages unreliable estimators. That much is standard.

Why: Because It Compounds

Take the same coin and the same probabilities, and change only the dynamics.

	multiplicative (×1.5 / ×0.6)	additive (+$0.50 / −$0.40)
mean	$13,150	$105.00
median	$0.52	$105.00
institutions needed for an estimate	tens of millions	one

In the additive world one institution tells you everything and the mean equals the median. In the multiplicative world they are separated by a factor of 25,000. The sampling problem is not a separate defect; it is the compounding, seen from the measurement side.

Which leaves one assumption carrying the whole letter, and it should be argued rather than assumed. Are the dynamics multiplicative? Three reasons to think so. More artificial intelligence in production means more at stake, so a period's consequence scales with current exposure rather than arriving as a fixed amount. Remediation cost scales with the deployed estate — rebuilding governance across forty models costs more than across four. And supervisory damage compounds, because a second finding lands harder than the first, the first having established a pattern.

If the dynamics were additive instead, everything here dissolves and one institution would suffice. That is the honest test, and stating it is what makes this falsifiable rather than merely persuasive.

Even Measured Perfectly, It Is Not Your Number

The measurement problem has a direction of travel. A supervisor can answer it by pooling across jurisdictions, lengthening the window and improving the data, and can reasonably say the estimates will get better.

The second argument has no such reply. Give the supervisor a flawless estimate of $13,150 and the individual institution still ends at $0.52. The objection is not that the estimate is imprecise. It is that the quantity is the wrong quantity for the reader it reaches.

> Which average is relevant depends on whose position you occupy, not on the quality of the data.

A supervisor holds a population, so the ensemble average is the right tool for a supervisor. A chief risk officer holds one institution on one path, so only the trajectory describes what will actually be experienced. Both answers are clear on their own. The failure is that a framework built for the first position is handed down to the second, where the same formula now answers a different question, and nothing on the page says so.

And in Artificial Intelligence, Neither Quantity Exists

The coin is the best case. Known multipliers, known probabilities, a stable process — and even there the two averages diverge by ten percentage points a round, and the cross-section is hopeless. An artificial intelligence programme has none of it. There is no repeated identical trial, no stable distribution, and a system that changes underneath you as models are retrained, data drifts and agents are granted new authority. There is no generating process to average over.

So the claim is not that we must choose between two computable averages. It is that neither exists to be chosen.

This error has a century-old name. Knight separated risk, where probabilities can be assigned, from uncertainty, where there is no basis for assigning them. Writing a confidence figure against a thing nobody can establish converts the second into the first by notation. Once it is a number it can be weighted, aggregated, reported and capitalised, and the whole apparatus of risk management applies to something it should not.

The harm is not that the formalism fails to help. A qualitative judgement — we cannot establish where this came from — is a stopping condition. A score of 0.8 is a passing mark. The notation converts something that should halt a decision into something that can be averaged away.

This cycle has already run twice. Value-at-risk assigned probabilities to tail events with no adequate frequency behind them, and the response after 2008 was largely more quantification. Structured credit ratings did the same with a letter grade that looked like measurement. In both cases the number was not a summary of evidence but a licence to proceed, and its main function was that it could be shown to someone afterwards.

The Same Error on Another Axis

Everything above concerns two directions, institutions and time. A chain of agents introduces a third — steps within a single decision — and the same arithmetic error appears there, which is worth seeing separately rather than folding into the argument above.

End-to-end reliability across a ten-step workflow is the product of the step reliabilities, not their average. Three chains, all reporting the same average.

chain	reported average	actual end to end
uniform, every step 0.95	0.95	59.9%
nine steps at 0.995, one at 0.545	0.95	52.1%
nine steps at 1.000, one at 0.500	0.95	50.0%

The same number on the report; ten percentage points apart in production. The average cannot see the weak link, and the product can see nothing else.

The same holds for a stack of controls. Five control points, each catching four problems in five, let 3 in 10,000 through if they fail independently. Give them a common blind spot of only 1% and the escape rate rises 32 times. A data foundation failure is invisible to every control downstream of it, which means the advertised effectiveness of the whole stack rests on an independence assumption that is neither stated nor tested.

What to Do Instead

A score is an ensemble construct. A gate is a trajectory construct. If the process were ergodic, scoring would be fine: you would be optimising a population average and it would converge on something real. Because it is not, admissibility conditions are not a stylistic preference. They are the correct object.

Is there a single authoritative source for this attribute — yes or no? A critical attribute that fails cannot be compensated by 99 that pass, because the compensation is an average, and averages are what this letter is about. The score keeps one narrow job afterwards: once admissibility has gated, a weighted score can rank remediation effort among the attributes that passed. Ordering work, rather than licensing action.

There is a harder question underneath, and it is institutional rather than technical. Who declares which source is authoritative? A chain of custody can be machine-verified; its root

is somebody accountable for saying that this system is master for this attribute. In most institutions that record does not exist, or it sits in a document nobody has tested against the systems. It is the situation Letter 4 described: when the data is questioned, there is nobody to ask.

The Regulatory Signal

On 7 October 2026 the Monetary Authority of Singapore issued its final Guidelines on Artificial Intelligence Risk Management. They take effect on 7 October 2027, with the lifecycle controls following by 7 October 2028. They apply to every financial institution and to every form of artificial intelligence.

The timetable is not the interesting part. The scope test is. The full guidelines apply where poor performance or unavailability of an institution's artificial intelligence is likely to have a material impact; where it is not, basic policies and procedures suffice. That is a gate — material or not — and it is doing the work that a maturity score is usually asked to do.

Read the clause once more, because it does not stop at the institution. The impact to be assessed is on the institution, its customers, "or other stakeholders, including other FIs". A firm is being asked to judge its own failure by what it would do to the rest of the column. That is the distinction this letter is about, arriving inside a supervisory expectation rather than a paper.

What the guidelines do not say is which average a reported figure is. An institution must assess materiality and apply proportionate controls, but nothing requires it to state whether the reliability figure it puts in front of its board is the population's or its own. The gate is the right instrument. The number passing through it is still unmarked.

The Financial Stability Board, which consulted on the same ground in June, published its final sound practices on 15 October 2026. Its fifth asks each institution to assess the materiality and risk of its artificial intelligence use cases "at the inception stage and thereafter" — again an assessment of one institution's own path, and again with nothing to say which average the resulting figures are. The Board also describes practices specific to agentic artificial intelligence as nascent. Singapore has said it will consult the sector on agentic artificial intelligence during 2027 — at which point the question of whether 95% is the average of the steps or the product of them stops being a matter of style.

Questions Worth Raising at Board Level

As in previous letters, these are framed so that a statement of policy will not satisfy them.

1. When management shows us a number, is it a column or a row?

A column is every institution in one year. A row is this institution across years. The two are not interchangeable, and almost nothing in a board pack says which one is on the page.

2. Was this figure computed where losses are shared, or where they are borne alone?

This is the same question in the form that works at every level — for a client reading a firm's figure, for a firm reading a sector figure, and for an institution reading a supervisor's.

3. When management reports 95% reliability for a chain of agents, is that the average of the steps or the product?

The difference is 95% and 60%. Most teams will not know which they are quoting. The moment they check, the argument has made its own case.

4. What would have to happen for this system to be stopped, and when was that last exercised?

A stop that has never been fired is a description, not a control. It is also the only thing that keeps a bad period from becoming the end of the row.

A Final Observation

None of this says the average is the wrong tool. It is the right tool for a population, and the supervisor who computes one is using it properly. The difficulty is that the number is true where it is computed and ceases to be true one level down, and nothing on the report marks the transition.

The institutions in next year's figure will be the ones that are still there. That is what the figure is for. It is not what you are.

The full argument, with the derivations, the exact distribution and the references, is set out in the working paper The Institution and the Supervisor Measure Different Averages of the Same Process, posted on SSRN on 28 September 2026: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=7536341

The coin-toss game in this letter is Ole Peters' canonical example; the paper carries the full citations, together with Kelly, Breiman and Knight.

About the Author

Reto Gruenenfelder advises boards and risk committees in financial services on getting AI right: programmes that deliver what the board approved, with their risks under control from the start. He spent almost 20 years with IBM, where his main responsibility was large bank transformations, finally as CTO Financial Services for IBM ASEAN. From 2019 to 2022 he was Technology Advisor to the Board of Bursa Malaysia.

He has worked independently since 2017, from Kuala Lumpur and Zurich, and publishes his experience as working papers. He is committed to making AI accessible to a wider audience.

Primary Source Documents

The following link directly to the original material referenced above.

→ Gruenenfelder: The Institution and the Supervisor Measure Different Averages of the Same Process — SSRN working paper, September 2026

→ Monetary Authority of Singapore: Guidelines on Artificial Intelligence Risk Management — 7 October 2026

→ Financial Stability Board: Sound Practices for Responsible Adoption of Artificial Intelligence (AI), Final report — 15 October 2026

→ Peters: The Ergodicity Problem in Economics — Nature Physics, 2019

Published for informational purposes only. Not legal, regulatory, or compliance advice.

Full disclaimer: itaiadvisory.com/disclaimer

© 2026 IT AI Advisory Sdn. Bhd. (202601023072). This letter may be shared freely in its original form.