

AI BENEFIT REALISATION

A joint CFO–CAIO framework for AI benefit realisation

Baselines, attribution, inference cost accounting and stage-gates that stop funding for pilots that miss their targets.

Terence Kok with Dr. Christina Chua

25%

of AI initiatives delivered the ROI their CEOs expected

Only 16% scaled enterprise-wide. IBM CEO Study, 2025, 2,000 chief executives.

Gartner expected at least 30% of generative AI projects to be abandoned after proof of concept by the end of 2025, citing escalating costs and unclear business value among the reasons.

Two executives have a claim on AI value. Each assumes the other is measuring it.

The CAIO

Owns the use case, the model, the data and the delivery team.

The CFO

Owns the budget, the benefits register and the numbers that go to the board.

The gap

Time released but never redeployed. A baseline taken in a quiet month. An inference bill growing faster than the saving.

76% of organisations had a Chief AI Officer in 2026, up from 26% a year earlier. IBM CEO Study, 2026.

Four controls the CFO and CAIO sign together

1 • Baseline

Signed by finance before the first production model call.

2 • Attribution

Method agreed at approval, before any results exist.

3 • Unit cost

Cost per outcome for every use case, reviewed monthly.

4 • Stage-gates

Scale, extend once, or stop, on a fixed date.

What a baseline record a CFO can sign contains

Outcome

Defined as finance would count it

Data source

From systems of record, with an owner

Period

At least one full business cycle

Variance

How much the metric moves on its own

Quality

Measured alongside speed

Cost to serve

Fully loaded, today

Target and kill threshold

Written before any results exist

Five attribution methods ranked by confidence

Randomised holdout

Highest confidence

Staggered rollout

High: difference-in-differences across waves

Matched comparison

Moderate

Trend-adjusted before and after

Low

Self-reported estimate

Not counted as realised benefit

Report each benefit with its method and grade. Never add different grades into one total.

5–30x

more tokens per task for agentic models than a standard chatbot

Gartner, March 2026: as token consumption rises faster than token costs fall, overall inference costs are expected to increase.

Track cost per outcome: inference, retrieval, evaluation, human review and platform share, tagged to the use case.

Four funding gates with kill criteria written in advance

Gate 0 • Fund the pilot

Baseline, method, target, kill threshold, cost ceiling, tagging

Gate 1 • Pilot review

Scale, extend once, or stop

Gate 2 • Scale

Unit benefit above unit cost at full volume

Gate 3 • Realisation review

Realised against forecast at 6 and 12 months

A portfolio where nothing has ever been stopped means the gates are not being used.

Five questions to ask

- 1. Which AI use cases had a signed baseline before they started?**
- 2. What attribution method sits behind each claimed benefit?**
- 3. What does one outcome cost today, including human review?**
- 4. When did we last stop funding an AI pilot?**
- 5. Does our AI inventory record the benefit case and the gate decisions?**

A gate that stops a weak pilot early is the cheapest control an AI programme will ever buy.

Read the full framework at terencekok.com



Terence Kok · AI & Innovation Strategist

SAVE · COMMENT · FOLLOW