94% model accuracy is not a business result. It's a capability spec.

Enterprise AI programmes across Asia are accelerating on every indicator except the one that matters: whether operations actually changed.

SWIPE →

Two measurement layers, constantly conflated

01 Model performance metrics (the technical layer)

02 Business performance baselines (the operational layer)

03 Attribution rules, defined before deployment

04 Governance that connects both layers

94%

accurate is a capability spec, not a business result

The business result is whether processing time, cost per transaction, or resolution rates actually moved.

Why Large Enterprises Are Measuring AI Incorrectly, 2026

What each layer actually tells you

MODEL LAYER	OPERATIONAL LAYER
Accuracy, F1 score	Processing time per transaction
Inference latency	Decision consistency rate
Model drift indicators	Cost per unit of output

Three structural reasons, not one bad habit

01 Programme sponsorship sits in tech, not operations

02 Vendors are contracted on model benchmarks alone

03 Baselines get installed with the model, not before it

"

The instrumentation required to capture pre-deployment baselines must be installed before the model, not alongside it. Concurrent installation makes a before-and-after comparison structurally impossible.

Terence Kok
Enterprise AI Strategist

Three layers where this catches up with you

01 Reporting — scale-up budgets approved without proven return

02 Risk — model drift degrades output before impact is visible

03 Governance — regulators want operational evidence, not model certificates

Extend measurement. Don't replace it.

Lock operational baselines before deployment. Define attribution rules with explicit time windows. Instrument the workflow, not just the model API. Connect drift alerts to operational review. Report operational impact to the board — model metrics stay in engineering dashboards.

Before or after?

Were your operational baselines locked before the model went into production — or only after, when it was too late to compare?

Full breakdown and the corrective framework:
terencekok.com/blog/why-large-enterprises-deploying-ai-scale-measuring-incorrectly

Model metrics say it works. Operational metrics say if it mattered.

Baseline before deployment, not after. It's the only sequencing that lets you tell the difference.

Terence Kok · Enterprise AI Strategist

SAVE · COMMENT · FOLLOW