

THE AI ARCHITECT LETTER · PLAYBOOK

The Operating-Model Multiplier

Two banks buy the same AI — one earns five percent, the other fifty. The difference is not the model. It is the operating model. A diagnostic for raising your multiplier.

"The model is the commodity. The operating model is the durable, compounding asset — the one advantage a competitor cannot buy **off the shelf**."

1**Rewire**

Redesign around the AI

2**Reallocate**

Capacity → P&L

3**Re-govern**

Enterprise layer

STEP 1 · SEE THE MECHANISM

Same capability, different return

The model is commoditizing — the same frontier models sit behind the same cloud interfaces, available to every competitor. So the edge is no longer what you buy. It is the machinery you build around it.

$$\text{Realized AI value} = \text{Raw Capability} \times \mathbf{M}$$

M = the operating-model multiplier

Bolt-on bank · $M < 1$

AI added as a feature on a legacy process. A reported **5–15%** local gain that leaks before the P&L. The operating model destroys value.

Rewired bank · $M = 2\text{--}3\times$

The operating model redesigned around the AI. A reported **20–50%** impact that compounds enterprise-wide.

M is a product, not a sum. It is set by three levers that **multiply** each other — so zero on any one lever is zero overall. A bank can rewire perfectly and still earn nothing if it never reclaims the freed capacity. The multiplier compounds only when all three clear one.

Why most banks get the small number. Buying a model is a procurement decision — fast, easy, photographable. Rewiring how thousands of people work is slow and political, so it gets skipped — and the gain dies at the same old queue. The capability rises; the operating model does not; the multiplier stays near one. That is why a reported ~73% of pilots stall.

STEP 2 · PULL THE LEVERS

The 3 multiplier levers

1

Rewire — redesign the process around the AI

Not bolt it on. The difference between the 5–15% number and the 20–50% number. Its signature is the spending mix — BCG's **70/20/10**.

Diagnostic: read the ratio and you can predict the return before any model is chosen.

2

Reallocate — convert freed capacity to the P&L

Freed time is not profit until reclaimed. By some estimates up to **two-thirds** leaks back into low-value work unless capacity is redeployed or removed.

Callback to the Visibility Trap — the productivity-not-profit leak, now as one lever of three.

3

Re-govern — governance as an enterprise layer

After SR 26-2, governance moves out of the model-risk silo onto the enterprise operating model — the rail that lets revenue and agentic AI scale, not the brake.

Callback to the 4 Stone Guardians — gate, guardrail, ledger, owner — placed at the enterprise layer.

The 70/20/10 signature

70 · people & process

20 · data & tech

10 · model

BCG's rule, corroborated by McKinsey QuantumBlack, Deloitte, and MIT Sloan. The banks that fail invert it — heavy on the model, light on the workflow.

STEP 3 · DIAGNOSE THE BROKEN LEVER

The broken-lever diagnostic

| Symptom | Broken lever | The fix |
|---|--------------|---|
| Big gain in the demo, little in the P&L | Rewire | Still bolted on — redesign the process around the AI; re-weight to 70/20/10 |
| Efficiency pilot succeeds, ROE doesn't move | Reallocate | Capacity leaking — redeploy or remove; make it a board-tracked number |
| High-value AI stalls in review | Re-govern | Stuck in a model-risk silo — move governance to the enterprise layer |
| All three above one, return compounds | Healthy M | Protect it; re-audit before any agentic scale-up |

The evidence, kept honest

70 / 20
/ 10

BCG rule (McKinsey QuantumBlack · Deloitte · MIT Sloan corroborate): ~70% of AI value from people & process, ~10% from the model. Winners out-execute, they do not out-purchase.

~73%

Reported share of banking AI pilots that stall before production — the operating model failed, not the model. (Figure varies across sources.)

17% →
29%

BCG projects agentic AI's share of AI value rising by 2028 (across industries) — agents are far less forgiving of a weak operating model, so the multiplier gap widens.

STEP 4 · RUN IT EVERY CYCLE

The Multiplier Audit

1

Measure M

For each scaled use case, compare the vendor's raw capability against the return that reached the P&L. The gap is your operating-model multiplier.

2

Find the broken lever

Is the gap from Rewire (still bolted on), Reallocate (capacity leaking), or Re-govern (stuck in a silo)?

3

Re-weight to 70/20/10

Move marginal spend off the model and onto process redesign and capability.

4

Lock the capacity

Make freed capacity a board-tracked number — redeployed or removed — not a slide.

The CFO test

"For our biggest AI use case, what is the gap between what the vendor's model can do and what actually reached the P&L — and which of the three levers is closing it?"

If the answer is "the model," you are out-purchasing. If it is "the operating model," you are building the multiplier.

THIS WEEK'S FRAMEWORK

Stop buying the model. Start building the multiplier.

aibusinessarchitect.ai/frameworks/ · [The AI Architect Letter](#)