ENTERPRISE AI PLAYBOOK · DEEP DIVE · PART 2 · DECIDE

Aligning AI Initiatives with Business Readiness

A readiness model for matching AI ambition to organizational capacity — before the budget is committed and the timeline is promised.

Readiness Gap = Ambition – Capacity

When the gap is wide, initiatives stall regardless of how good the technology is. You close it two ways: fund the capacity, or shrink the ambition. Pretending the gap isn't there is not one of them.

- **Dimension 1 · Strategic Fit — tied to an outcome, or to FOMO?**
- **Dimension 2 · Data Foundation — can you actually feed it?**
- **Dimension 3 · Platform & Skills — can you build it and run it?**
- **Dimension 4 · Process Fit — can the workflow absorb it?**
- **Dimension 5 · Risk & Governance — can you stand behind the output?**

Tom Ward
Enterprise Architect · Cloud & AI

linkedin.com/in/tomward
July 2026

DIMENSION 1

Strategic Fit

"Tied to an outcome, or to FOMO?"

The fastest way to waste an AI budget is to start with the technology and go looking for a problem. Strategic fit means every initiative traces to a business outcome someone in the business — not IT — is willing to be measured on. "We need an AI strategy" is not a business outcome. "Cut claims-processing cycle time by 30%" is.

The architect's job here is unglamorous and decisive: force the outcome to be named, quantified, and sponsored before capacity is committed. An initiative without an executive sponsor who will defend its budget in the next downturn is not ready — it is a science project with good PR.

KEY DECISIONS

- What measurable outcome does this initiative move, and who owns that number?
- Is the sponsor senior enough to hold funding through the trough of disillusionment?
- What are the kill criteria — the conditions under which you stop?

WATCH OUT FOR

- × Board-driven "do something with AI" mandates with no target attached.
- × Initiatives justified by the technology's novelty rather than the outcome's value.
- × Sponsors who want the credit but won't defend the budget line.

WHEN IT FAILS · IBM WATSON FOR ONCOLOGY

Watson was positioned as a cancer-treatment adviser on the strength of its technology, not a clinical outcome any oncology department owned. MD Anderson halted its project in 2016 after roughly **$62M** with no clinically usable system; IBM divested Watson Health's assets in 2022. **The outcome was never named by the people who'd be measured on it.**

Source: STAT News (2018)

DIMENSION 2

Data Foundation

"Can you actually feed it?"

This is the dimension that quietly kills more AI programs than any other, because the demo runs on a clean sample and production runs on the real thing. Before you promise an outcome, verify that the data the model needs is accessible, governed, of sufficient quality, and legal to use for this purpose. All four. Missing any one turns a six-week build into a six-month data-remediation project nobody scoped.

Readiness here is rarely all-or-nothing. The useful question is not "is our data ready?" but "is the data for *this specific use case* ready, and if not, what is the shortest path to making it so?" That reframing turns a paralyzing enterprise-data problem into a scoped, fundable task.

KEY DECISIONS

- Is the required data accessible today, or does it need integration work first?
- Who owns its quality, and what is the freshness SLA the use case needs?
- Do usage rights and privacy constraints permit this application of the data?

WATCH OUT FOR

- × Demos on curated samples that hide the real data's mess.
- × "We have lots of data" mistaken for "we have usable data for this."
- × Privacy or contractual limits discovered after the build, not before.

WHEN IT FAILS · AMAZON'S RECRUITING ENGINE

Amazon trained a résumé screener on ten years of its own hiring data — plentiful, accessible, and encoding a decade of male-skewed decisions. The model learned to penalize résumés containing "women's." The project was disbanded by 2017. **"We have lots of data" was true; "we have the right data for this" was not.**

Source: Reuters (Oct 2018)

DIMENSION 3

Platform & Skills

"Can you build it and run it?"

Building a proof of concept and running a production capability are different sports, and the gap between them is where readiness is won or lost. The platform question is whether you have the integration surface, the deployment pipeline, and the operational tooling to run an AI system at Tier-1 standards — not whether a talented engineer can wire up a demo.

Skills readiness is honest headcount math. You need people who can build, but more scarcely, people who can *operate* — monitor, evaluate, and improve a non-deterministic system over time. If that capability doesn't exist in-house and isn't budgeted, the initiative's real timeline includes a hiring or partnering plan, and the plan is part of readiness.

KEY DECISIONS

- Does the platform support deploy, monitor, evaluate, and roll back — or just build?
- Build in-house, buy a platform, or partner? Match the choice to your skills reality.
- Who operates this on day 90, and are they hired yet?

WATCH OUT FOR

- × Confusing a heroic prototype with a repeatable delivery capability.
- × Buying a platform to compensate for skills you still need to build.
- × No plan for who runs it once the launch team moves on.

WHEN IT FAILS · ZILLOW OFFERS

Zillow scaled algorithmic home-buying faster than it could operate the model behind it. When pricing drifted against a moving market, the evaluation loop to catch it wasn't there. The unit wound down in Nov 2021: over **$540M** in write-downs, 25% of staff cut. **Building the model was never the hard part — operating it was.**

Source: CNBC on Zillow Q3 2021 results

DIMENSION 4

Process Fit

"Can the workflow absorb it?"

AI amplifies a process; it does not fix a broken one. Dropping a model into a workflow that is unstable, undocumented, or riddled with exceptions produces automated chaos faster than manual chaos. Process readiness asks whether the target workflow is well-understood and stable enough that inserting AI improves it rather than obscuring its problems.

It also asks a design question most teams skip: where does the human stay in the loop? The right answer depends on the error tolerance you named in Dimension 1. High-stakes decisions need a review step; low-stakes, high-volume tasks can run with lighter oversight. Designing that boundary deliberately — rather than by accident — is a readiness marker.

KEY DECISIONS

- Is the target process stable and documented, or still in flux?
- Where does a human review output, and what triggers escalation?
- How do handoffs between AI and people actually work day to day?

WATCH OUT FOR

- × Automating a process nobody has mapped — you inherit every hidden exception.
- × All-or-nothing automation where a human-in-the-loop step was the safer design.
- × Ignoring the frontline workers whose job the workflow change reshapes.

WHEN IT FAILS · AIR CANADA'S SUPPORT CHATBOT

A support chatbot invented a bereavement-fare policy that didn't exist. In 2024 the BC Civil Resolution Tribunal held the airline liable for negligent misrepresentation, rejecting its claim that the bot was a separate legal entity. **With no review step between output and customer, the workflow shipped the error straight out — and the liability straight back.**

Source: Forbes (Feb 2024) on Moffatt v. Air Canada

DIMENSION 5 Risk & Governance

"Can you stand behind the output?"

Every AI initiative carries risk the organization must be willing to own: wrong answers, biased outputs, leaked data, regulatory exposure. Readiness here is not "have we eliminated risk" — you can't — but "have we set our risk appetite deliberately, mapped the compliance obligations, and named who is accountable when something goes wrong." In regulated industries, this dimension often gates the others.

The architect's leverage is to make this concrete and early. Map each initiative to its regulatory context, decide the risk tier before building, and give it a named accountable owner. Governance treated as a launch-gate checkbox fails; governance treated as a design input — the way NIST's AI RMF frames it — turns risk from a blocker into a set of requirements you can actually build against.

KEY DECISIONS

- What is the risk tier of this use case, and does our appetite cover it?
- Which regulations apply, and are the obligations mapped to controls?
- Who is accountable — by name — for the system's outputs?

WATCH OUT FOR

- × Discovering the compliance requirement at the launch review.
- × Diffuse accountability — "the AI decided" is not an answer a regulator accepts.
- × A risk appetite that exists on a slide but not in the actual go/no-go decision.

WHEN IT FAILS · APPLE CARD CREDIT LIMITS

In 2019 the Apple Card drew viral complaints that it granted men far higher limits than their spouses on shared finances. New York's DFS later found no unlawful discrimination — but faulted the program's transparency and customer service. **The lesson is not that the model was biased — it is that no one was positioned to stand behind the output when asked to.**

Source: TechCrunch (Mar 2021) on the NY DFS findings

SPANNING Change Capacity

"The multiplier on all five dimensions"

An organization can be ready on all five dimensions and still fail on the one nobody budgets for: its capacity to absorb change. A workforce mid-reorg, already digesting three transformations, has little bandwidth left for a fourth — no matter how sound the initiative. Adoption depends on trust and training; sponsorship depends on stamina that outlasts the hype cycle. Treat change capacity as the multiplier it is: strong readiness scores times zero change capacity still equals a stalled program.

The Readiness Gap, in Public

Five documented failures. In each, the gap was on the readiness side — and visible before launch.

IBM Watson for Oncology [1] 2013–2022

STRATEGIC FIT

Sold as a cancer-treatment adviser on the strength of a Jeopardy! win. Training leaned on synthetic cases rather than real patient outcomes; internal documents cited "unsafe and incorrect" recommendations. MD Anderson halted its project in 2016 after roughly **$62M** with no clinically usable system. IBM divested Watson Health in 2022.

Readiness signal missed: no clinical outcome owned by the department being sold the tool. Ambition was set by the demo, not by a number anyone agreed to be measured on.

Amazon Recruiting Engine [2] 2014–2018

DATA FOUNDATION

A résumé-ranking model trained on a decade of Amazon's own hiring decisions learned the pattern in that data: it penalized résumés containing "women's" and downgraded graduates of two all-women's colleges. Editing out those terms couldn't guarantee the model wouldn't find other proxies. The project was disbanded by 2017.

Readiness signal missed: plentiful, accessible, high-quality data that was still the wrong data for this purpose. Volume is not fitness.

Zillow Offers [3] 2021

PLATFORM & SKILLS

An algorithmic home-buying arm scaled to thousands of purchases per quarter. CEO Rich Barton conceded "the unpredictability in forecasting home prices far exceeds what we anticipated" — the loop to catch that drift wasn't there at scale. Wound down Nov 2021: over **$540M** in write-downs, 25% of staff cut.

Readiness signal missed: the capability to *operate* a non-deterministic system — monitor, evaluate, roll back — never scaled with the capability to build it.

Air Canada Chatbot [4] 2022–2024

PROCESS FIT

A website chatbot told a grieving customer he could claim a bereavement fare retroactively — contradicting the airline's own linked policy page. Air Canada argued the bot was "a separate legal entity" answerable for its own statements. The tribunal rejected that and found negligent misrepresentation.

Readiness signal missed: no review step between generated text and a binding customer commitment. The human-in-the-loop boundary was set by default, not by design.

Apple Card Credit Limits [5] 2019

RISK & GOVERNANCE

Viral complaints alleged the model gave women far lower limits than spouses with shared assets. New York's DFS reviewed ~400,000 applications and found *no* unlawful discrimination — decisions were explainable and policy-consistent. But it faulted the program's transparency and customer service for undermining trust.

Readiness signal missed: a lawful model is not a defensible one. Regulators could get the underwriting rationale; the affected customers could not. Explainability that only satisfies an auditor still leaves you unable to answer the person asking.

SOURCES: [1] STAT News (2018) on Watson's "unsafe and incorrect" recommendations [2] Reuters, J. Dastin (Oct 2018) on Amazon scrapping its AI recruiting tool [3] CNBC (Nov 2021) on Zillow's Q3 results & Offers wind-down [4] Forbes (Feb 2024) — ruling at Moffatt v. Air Canada, 2024 BCCRT 149 [5] TechCrunch (Mar 2021) — primary: NY DFS press release, Mar 23, 2021

The pattern: none were model-quality failures. Each was a readiness gap this assessment would have surfaced before the budget was committed.

The Readiness Scorecard

Score each dimension 0–4 for a specific initiative (not your organization in the abstract): 0 = absent, 2 = partial, 4 = solid. Then set that total against your honest change capacity. The point is not a high score — it is a deliberate decision: proceed, fund the gap, or shrink the ambition until it fits.

1 · Strategic Fit	Named, measurable outcome with a sponsor who will defend the budget.	☐☐☐☐
2 · Data Foundation	Data for this use case is accessible, governed, quality, and legal to use.	☐☐☐☐
3 · Platform & Skills	You can build, deploy, operate, and improve it — with people who exist.	☐☐☐☐
4 · Process Fit	Target workflow is stable, mapped, with a deliberate human-in-loop design.	☐☐☐☐
5 · Risk & Governance	Risk tier set, compliance mapped, a named accountable owner.	☐☐☐☐
× Change Capacity	Bandwidth, trust, training, and sponsor stamina to absorb the change.	☐☐☐☐

0–9
Not ready — shrink the ambition or fund the gaps first

10–15
Conditional — proceed on the strong dimensions, close the weak ones

16–20
Ready — align capacity and go

One discipline separates organizations that scale AI from those that accumulate pilots: they run this assessment *before* committing, and they let a low score change the plan — a smaller scope, a phased rollout, a data-remediation step first — rather than powering ahead and discovering the gap in production.

Going Deeper: References

Sources worth your time, spanning strategy, data, risk, and change.

McKinsey — The State of AI
The adoption benchmarks to calibrate your ambition against what peers actually achieve.
mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

MIT Sloan Management Review — AI & Business Strategy
Research on tying AI initiatives to strategy — the core of Dimension 1.
sloanreview.mit.edu/big-ideas/artificial-intelligence-business-strategy

Deloitte — State of AI in the Enterprise
Enterprise readiness survey: what separates high-outcome adopters from the rest.
deloitte.com — State of AI in the Enterprise

NIST AI Risk Management Framework
The governance vocabulary for Dimension 5 — Govern, Map, Measure, Manage.
nist.gov/itl/ai-risk-management-framework

Prosci ADKAR — Change Management
A practical model for the change-capacity multiplier most AI plans ignore.
prosci.com/methodology/adkar

WEF — Empowering AI Leadership
A board-level toolkit for the executive-sponsorship and oversight questions.
weforum.org — Empowering AI Leadership (PDF)

OECD.AI Policy Observatory
Cross-jurisdiction policy and regulation context for mapping compliance early.
oecd.ai/en

About the Author

Tom Ward is an Enterprise Architect with 20+ years across Fortune 500 financial services, retail, and government — including Wells Fargo and Bank of America — specializing in EA governance, reference architectures, hybrid cloud, and legacy modernization, with recent hands-on GenAI and mobile delivery work.

Let's connect: linkedin.com/in/tomward