Reading the Story

Business cases, vendor claims, and the five ways enterprise AI predictably fails

Part One gave you the vocabulary. This guide gives you comprehension — because enterprise AI arrives at your level as claims: a business case, a vendor deck, a programme update, a benchmark from a consultancy. Each is a story told by someone with a stake in how it lands, in a field where verification is unusually hard and enthusiasm is unusually high.

By the end you will read AI business cases the way you learned to read forecasts — assumptions before outputs — recognise the five failure patterns early enough to intervene, and understand what is happening to your function's position while all this unfolds.

A reading note: several of these patterns will describe initiatives your organisation is currently running, possibly including one of yours. That is the normal state of an immature field. Seeing them clearly is what lets you be the person who names them constructively rather than the person surprised by them later.

John Obidipe · catalystgrowthcoach.co.uk

01

Reading an AI business case

Everything from the finance series applies unchanged — assumptions before outputs — with three AI-specific pressure points:

- **The productivity assumption.** Most AI cases rest on time saved per person per week, multiplied by a lot of people. Two questions dismantle a weak one: was that measured on our work or quoted from a vendor benchmark, and does the saved time convert into anything the P&L can see? Twenty minutes a day returned to five hundred people is either a substantial capacity gain or nothing at all, depending entirely on whether the organisation does something deliberate with it.
- **The accuracy assumption.** Pilots report accuracy on curated data with attentive users. Production has messy data, edge cases, and people who stop checking once it seems to work. Ask what the error rate was on production-like data, and what the checking cost is at full volume — because that cost is permanent and usually missing from the case.
- **The context assumption.** Most enterprise AI cases quietly assume the organisation's own information is available, current and correct, because that is what the tool will answer from. This is now recognised as the discipline that decides whether these systems work at all, and it is where most disappointing initiatives actually fail. "What does this depend on knowing about us, and is that information in good enough shape?" is often the question that reveals the real project underneath the AI project.
- **The adoption assumption.** AI cases assume people use the tool as intended. The technology series' demo-to-production gap applies with force here, because AI adoption also requires behaviour change and trust. "What is the plan for the people, and who owns it?" exposes cases where the business change was assumed rather than resourced.

And the run-cost tail, from the technology series: AI adds permanent cost that scales with usage. "What does this cost at full adoption, and what are we switching off to fund it?" remains the sharpest single question in any technology case, AI included.

02

Reading vendor claims

- **The Tuesday test, enterprise edition.** "Show it working on our data, with our messiness, end to end." Applies unchanged from a five-person business to a five-thousand-person one, and remains the fastest way to separate capability from choreography.
- **Ask for the error workflow, not the accuracy number.** Serious vendors describe detection, escalation, and correction. "It's highly accurate" without a described failure path means nobody has thought about failure, which means you will.
- **Ask which model is underneath, and what happens when it changes.** Model providers update and deprecate; behaviour shifts. A vendor who has thought about version management has thought about running a real service.
- **Ask about integration, and name the standard.** Integration complexity, not model capability, has been the real brake on enterprise AI. Ask whether the vendor supports MCP, the open standard the major providers converged on during 2026. A yes changes the cost and timeline of anything described as a custom build. Ask in the same breath who reviews what a connection can reach, because over-permissioned connections are where the security attention has moved.
- **Treat benchmarks as marketing until proven otherwise.** Published productivity figures come overwhelmingly from parties with an interest in them. Your own two-week measurement on your own work is worth more than any published percentage, and takes less time to obtain than most procurement processes.

– **Reference calls, not logo slides.** Ask specifically for a reference at your scale who has been live for more than a year — the difference between 'we bought it' and 'we run it' is where the useful information lives.

03

Five patterns of enterprise AI failure

– **The proof-of-concept graveyard.** Dozens of pilots, few in production. The most widespread enterprise AI pattern by a distance. Cause: pilots approved without promotion criteria — nobody agreed in advance what result would graduate them, so nothing ever does, and the portfolio accumulates unresolved experiments that look like activity.

– **The productivity claim nobody can find.** The programme reports hours saved; the P&L shows no change and headcount is identical. Not necessarily fraud — usually the saved time was real and diffuse, returned to individuals in twenty-minute slices and quietly reabsorbed. Time saved is only a benefit when something specific is done with it, and that conversion is a leadership act, not a technology outcome.

– **Oversight that isn't.** Human-in-the-loop on paper; in practice one reviewer approving hundreds of items with no realistic capacity to assess them. The control exists in the documentation and not in the world — exactly the kind of finding that reopens annually in internal audit, and exactly the risk regulators are most alert to.

– **Shadow AI outpacing governance.** The organisation debates policy while the workforce adopts tools weekly. By the time a framework lands, practice has moved on and the framework describes a world that no longer exists. Governance that cannot keep pace is governance that gets routed around.

– **The capability that was never a capability.** One enthusiastic team builds something impressive that depends entirely on them: undocumented prompts, unowned data pipelines, no support model. It works beautifully until they move on. The key-person system, in its newest costume.

– **The agent that was an automation.** 'Agent' is 2026's most inflated word, and a substantial share of what arrives labelled agentic is a scripted workflow at agentic pricing. The governance consequence matters more than the money: a genuine agent chooses its next action and therefore needs scope limits, reversibility and audit; an automation does not. Approving one under the other's regime wastes control effort in one direction and creates unsupervised action in the other.

The shared root, as ever: none is a technology failure. Promotion criteria, benefit ownership, realistic control design, pace of decision-making, and institutional knowledge — governance and leadership failures wearing an AI badge. Which is why more AI investment resolves none of them.

04

Reading what AI is doing to your function

Two things are happening simultaneously, and directors who read only one of them get caught out.

The first is exposure. Functions whose work is substantially drafting, summarising, first-pass analysis or routine response are being examined for AI-driven efficiency, and 'what could AI do here?' is now a standing question in cost reviews. If that describes your function, the initiative you sponsor yourself is considerably more comfortable than the one an outside consultancy proposes about you.

The second is opportunity, and it is larger. AI capability is currently scarce, visible, and valued at executive level — and almost nobody in most organisations has both business judgement and enough AI fluency to lead sensibly. That combination is what executive committees are actively short of. A director who can say 'here is where we tested it, here is the measured result, here is what we're doing next and what we stopped' is offering something genuinely rare: evidence, in a conversation currently running on assertion.

05

An exercise — fifteen minutes

- Take the most recent AI business case you've seen. Find its productivity assumption and ask where the number came from.
- List your function's AI pilots. For each: what result would promote it, and who decides?
- Check one human-in-the-loop control honestly — items per reviewer, per hour.
- Identify which of the five patterns your organisation is running. Most run at least two.
- Ask your team what they use unofficially. Treat the answer as intelligence, not misconduct.

Where Part Three takes you

You can now read the cases, the claims, and the patterns. Part Three, "Asking the Questions," is about the seats you will occupy: sponsoring an AI initiative that survives contact with production, governing agentic AI without either paralysis or naivety, and leading a function whose people are anxious, curious, and already experimenting.

Because the destination in this territory is unusual: there is no expert to defer to. The vendors are selling, the consultancies are generalising, and the technologists own capability rather than consequence. The questions — and the judgement about what is worth doing — are genuinely yours.

John Obidipe is a business transformation specialist and Gallup-Certified CliftonStrengths coach. He works with founders and leadership teams on the structural side of growth — the part the numbers can only hint at. · coach@catalystgrowthcoach.co.uk · catalystgrowthcoach.co.uk