Architecture at the Table



A scorecard for architects and technology leaders

AI Production Readiness Scorecard

Most enterprise AI pilots do not fail because the model is wrong. They fail because five architecture decisions were never made. Here are the five, and a scorecard to use before your next pilot.

Why pilots stall

The demo worked. The board loved it. Six months later, it is still a demo.

A pilot is designed to answer one question: can this work? So we pick clean data, a supportive team, twenty carefully chosen test questions, and no security review, because it is "just a pilot". And it works. We removed everything that makes production hard.

Then someone asks the real questions. Can it use live customer data? Who approves its answers? What happens when it is wrong? What will it cost at ten thousand users? Who supports it at two in the morning?

The reframe

The demo is maybe ten per cent of the work. The other ninety per cent is architecture, and it comes down to five decisions.

Context: Gartner (July 2024) predicted at least 30% of generative AI projects would be abandoned after proof of concept by the end of 2025. MIT NANDA, "The GenAI Divide: State of AI in Business 2025", found most generative AI pilots show no measurable P&L impact.

The five decisions

1 Data readiness is the project

In the pilot, someone exported a spreadsheet, cleaned it by hand and loaded it in. In production, that data lives in six systems, changes every hour, some of it is personal, some of it is wrong, and nobody owns it.

- Where does the data really live?
- How fresh does it need to be?
- Who is allowed to see what?
- How do we keep access permissions intact when the AI retrieves a document, so it never shows a junior employee the board papers?

The decision
Design the data pipeline, the access model and the ownership before you choose a model. The model is the easy part.

2 Build a platform, not a collection of pilots

Marketing builds a pilot. So do finance and customer service, each with its own model, vector database, prompts and logging, or no logging at all. Within a year: seven AI tools, seven sets of risks, seven invoices, and no way to see what any of them is doing.

- One gateway that every application uses to reach the models
- Shared retrieval, guardrails, monitoring and cost tracking
- Teams build their own use cases on common foundations

The decision
Your second use case should take half the time of your first. If it does not, you are building pilots, not a platform.

3 Decide what good looks like before you build

The same question can get a slightly different answer every time, so "it looked good in the demo" is not a quality standard. The pilot passes twenty hand-picked questions; in week one a real customer asks question twenty-one, and gets a confident, completely wrong answer.

- Hundreds of real questions with agreed good answers, including the awkward ones and edge cases
- Every change to a prompt, model or data scored against that set before it goes live

The decision
Build an evaluation set and a pass mark, so you find question twenty-one before your customer does.

4 Design governance in, not bolted on

Many programmes treat governance as the final gate: the team builds for six months, risk and compliance see it for the first time, and it stalls for another six.

- Tier use cases by risk, early: meeting-notes summaries are low risk; loans, hiring or medical advice are high risk
- Match controls to the tier: a human in the loop for high-risk decisions, logging of prompts and responses, clear labelling, and a named person accountable
- A fast lane for low risk, a proper review for high risk

The decision
Governance stops being the brake and becomes the route to production.

5 Who owns this on day two?

The pilot team is a few enthusiastic people and a vendor. When the pilot ends, they move on, and nobody owns the thing.

- Somebody watches quality every week
- Somebody owns the budget, because AI costs grow with usage, not with licences
- Somebody handles incidents, re-tunes when the data changes, and decides when to retire it

The decision
If you cannot name those people before go-live, you are not ready to go live.

The scorecard

Before you approve your next AI pilot, ask these five questions and score yourself honestly.

Question	Yes	No	What would make it a yes?
1. Is the data pipeline and access model designed for production, not just for the demo?	Yes	No	
2. Is it built on a shared platform, or is it another one-off?	Yes	No	
3. Do we have an evaluation set, and a pass mark?	Yes	No	
4. What is the risk tier, and are the controls designed in?	Yes	No	
5. Can we name the five owners?	Yes	No	

The rule. Fewer than three yeses? Do not start the pilot yet. Fix the gaps first. It will be faster in the end.

The five owners

Before any AI system goes into production, write down five names.

Role	Name	Agreed on
Business owner		
Technical owner		
Data owner		
Risk owner		
Budget owner		

Regulation: EU AI Act, Regulation (EU) 2024/1689. UK Government, "A pro-innovation approach to AI regulation" (2023 to 2024).

Architecture at the Table

A weekly show on enterprise AI, architecture and technology strategy, for architects, CTOs and the business leaders who fund them. New episodes every Tuesday and Thursday.

Watch the episode Your AI Pilot Isn't Failing Because of the Model · youtu.be/BjhQjh_b5Co

Subscribe youtube.com/@ArchitectureAtTheTable

More guides and case studies architecture.vinava.co.uk

Free to share inside your organisation. Please keep this page attached.