Architecture at the Table



Case study pack

Seven AI Decisions, Worked Through

Worked examples from Architecture at the Table: AI agents, AI costs, governance, shadow AI, legacy systems, board decisions and how architects get heard. Written to share with your team.

architecture.vinava.co.uk

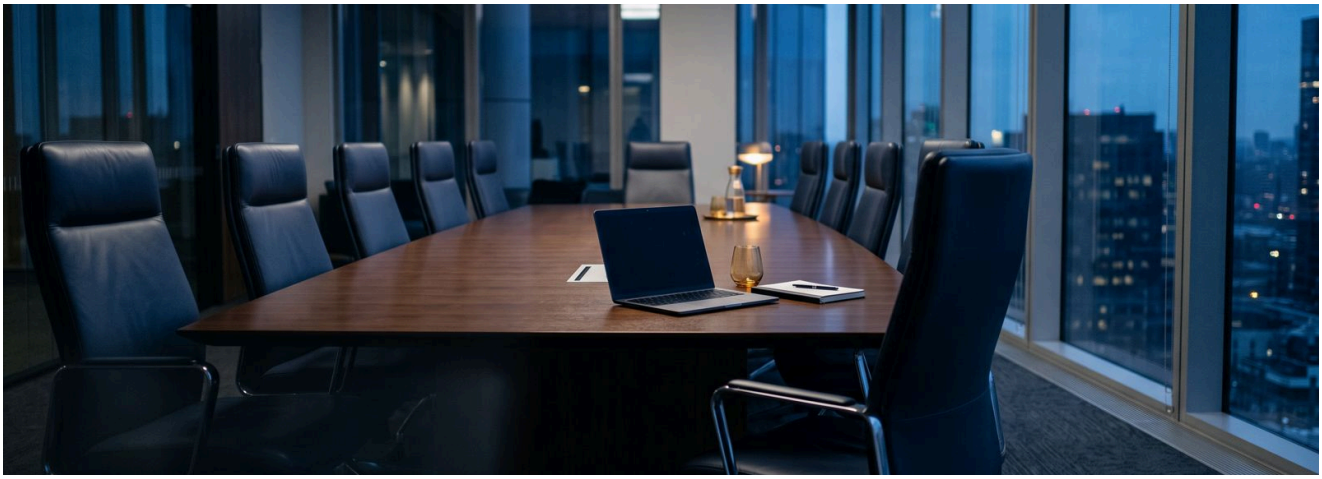
Contents

Each case study is an illustrative scenario, not a named client. Each one ends with a question to discuss with your team.

The refund agent	AI agents · episode 4
The contact centre assistant	AI costs and FinOps · episode 5
Two teams, one week	Architecture governance · episode 6
The finance team and shadow AI	Shadow AI · episode 7
The claims system	Legacy modernisation · episode 8
Two AI plans, one board meeting	AI strategy and boards · episode 9
The same message, twice	Influence and communication · episode 10

The refund agent

Designing an AI agent that can act, safely



The situation

A retailer wants an AI agent that handles customer refunds. Agents can now *act*: refund customers, change prices, approve invoices. So the real question is not whether it works in a demo. It is who said yes when it gets something wrong at 2am.

The design

- **Its own identity.** Every refund carries the name of the customer service colleague the agent is working for.
- **Clear permissions.** It can read orders and policies freely. It can issue refunds up to £50 on its own.
- **A human above the line.** Anything above £50, it drafts, and a person approves with one click.
- **Hard limits.** It can never touch prices, and it can never refund the same order twice.
- **A flight recorder.** Every decision is recorded, with the policy it relied on.
- **A kill switch.** There is a daily cap on total refunds. If the cap is hit, the agent pauses itself and alerts a manager.
- **A named owner.** The head of customer service owns it. Their name is on it.

The lesson

Nothing in that design is exotic technology. It is ordinary architecture discipline, applied to a new kind of worker. That is why architects need to be in the room before agents go live, not after the first incident.

Discuss with your team: Which of your systems could act on its own today, and whose name is on it?

The contact centre assistant

Designing AI for cost, on purpose



The situation

A contact centre assistant helps agents answer customer emails. It works, and the bill grows with every success.

The version that surprises the finance director

- Every email goes to the most powerful model.
- The full customer history and the entire returns policy are attached every time.
- The assistant drafts, checks its own draft and redrafts, three or four times.
- Nothing is cached, even though most emails ask the same ten questions.
- The cost lands in one central IT budget, so the contact centre never sees it.

The redesign

- A small model sorts each email by type.
- Common types are answered from a cache of approved responses.
- Only genuinely unusual emails go to the powerful model, with just the relevant paragraphs of policy.
- One draft and one check, then it hands over to the human.
- The contact centre owns the budget, measured as **cost per email resolved**.

The lesson

Same capability. Same quality, proven on the evaluation set. A fraction of the calls. None of it needed a new vendor or a new contract. It needed someone to design for cost from the beginning.

Discuss with your team: Do you know your cost per outcome for each AI use case, and who sees that number?

Two teams, one week

What replaces the architecture review board



Team A: a new customer API

They use the standard API pattern, on the AI platform, with approved data services. Their pipeline runs the automated checks on every change, and everything passes. They write a one-page decision record about the one small choice they made, and they ship. No meeting. No queue. Same week.

Team B: AI that recommends credit limits

That is new, and it affects customers directly, so it is tier three. An architect has been working with the team since day one, so there are no surprises. The review happens within two days, with a risk specialist and a data expert in the room. They agree three conditions:

- human approval above a threshold
- a monthly bias check
- a named owner

Team B ships a week later, with conditions everyone understands.

The lesson

Two very different risks. Two proportionate responses. And neither team had any reason to go around the architects.

Discuss with your team: Which of your recent changes needed a meeting, and which only needed an automated check?

The finance team and shadow AI

Turning a policy breach into a better tool



The situation

An AI amnesty reveals that the finance team has been using a consumer AI tool to summarise supplier contracts. Confidential data, in an unapproved tool.

The old response

A stern email and a ban. The usage goes underground, and the risk stays.

The better response

The architecture team sits down with finance and asks: what problem were you solving? They were spending two days a month reading renewal terms. Within a few weeks there is a proper contract summary assistant on the approved platform:

- Confidential data stays inside the organisation.
- Every summary links back to the clause it came from, so it can be checked.
- It works better than the shadow tool did.

The lesson

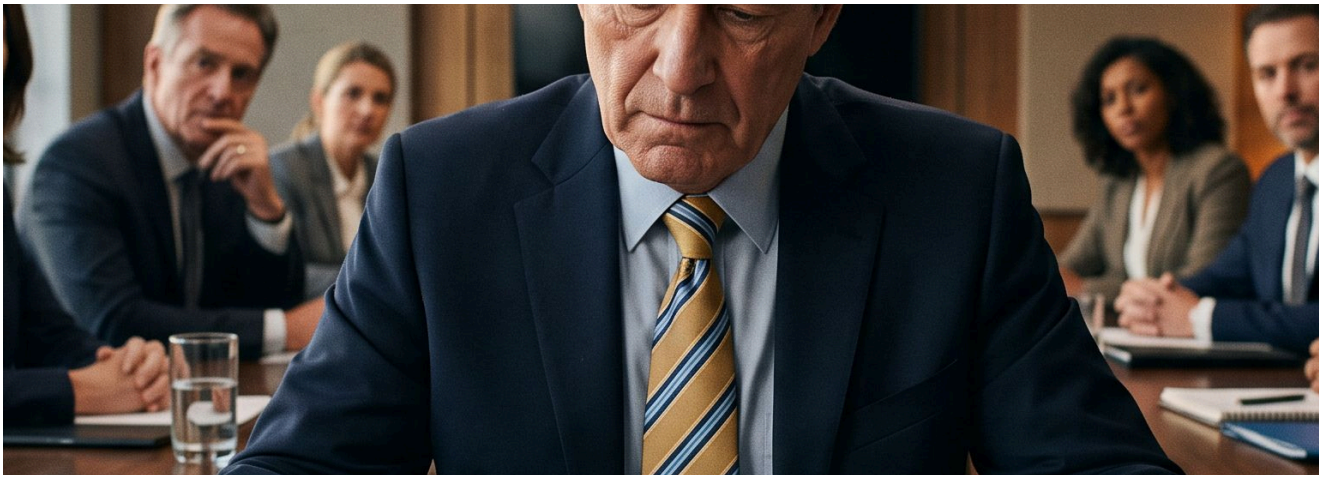
Same need. Safer answer. And a team that will come to you next time, instead of going around you.

Discuss with your team: What problem is your most popular unapproved AI tool solving, and can you solve it better?

Legacy modernisation · Illustrative scenario, drawn from episode 8

The claims system

Getting AI value out of a 30-year-old core system



The situation

An insurer wants AI to triage incoming claims. The claims data lives in a thirty-year-old core system, updated overnight.

Stage 1: understand it

AI reads the claims code and drafts the business rules. The two claims system experts review them over a few weeks.

Stage 2: wrap it

New claims are published as events into the data layer the moment they arrive, and an API is built to update a claim's priority. Within months, the triage AI is live, reading today's claims and setting priorities through a locked front door. Nobody touched the core.

Stage 3: replace it, piece by piece

The triage component becomes the first piece of the new claims platform. The next piece is chosen by the same value-and-pain score.

The lesson

The business saw AI value in months, and the modernisation pays for itself as it goes.

Discuss with your team: Which legacy system blocks your AI plans, and what would it take to put a safe front door on it?

Two AI plans, one board meeting

The questions that separate a strong plan from a demo



Plan A

Transform customer service with a leading AI vendor. A strong demo. But when the questions come, the answers are vague:

- The outcome is "better experiences".
- The cost is "a licence".
- Quality is proven by the demo.
- Ownership sits with IT.
- Nobody has thought about stopping.

Plan B

Smaller. Cut the time to answer supplier invoice queries from three days to one. It shows:

- a cost per query, compared with today
- a test set of 500 real queries with a pass mark
- a named owner in finance
- a stop rule: if the time saving is not visible within three months, it stops

The decision

Plan B gets funded. Not because it is less ambitious, but because it is the one that can prove it worked. Once it does, it earns the right to grow.

Discuss with your team: Which of your funded AI plans has a stop rule?

The same message, twice

Why leaders ignore architects, and how to fix it



Version one

The architect sends a twelve-page document explaining that the company's AI pilots lack a shared platform, with a seven-layer diagram and a list of risks. It is accurate. It is ignored.

Version two

The architect asks for fifteen minutes and brings one page:

- **Decision needed:** approve a shared AI platform for the next three AI use cases.
- **Options:** keep funding each pilot separately, or fund a platform once.
- **Recommendation:** fund the platform. Each new use case gets faster and cheaper, and AI spend becomes visible by team.
- **If we do nothing:** every new AI project keeps costing as much as the first.

One picture. One ask. And the architect says: "Yes, we can start next month, if we pick the first use case together."

The lesson

Same architect. Same facts. A completely different outcome.

Discuss with your team: What is the one decision you need from leadership this month, on one page?

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 Architecture at the Table: all episodes · youtube.com/@ArchitectureAtTheTable

Subscribe youtube.com/@ArchitectureAtTheTable

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

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