

Shipped a financial-services team's first production AI workload in 3 days — with evaluation, guardrails and governance built in from the first commit

US Financial Services — 3 days

We ran a 3-day intensive that took a US financial-services team from AI experimentation to their first production-ready workload — delivered with an evaluation harness, guardrails and infrastructure-as-code, and aligned with the governance and model-risk expectations a regulated financial business must meet.

THE CHALLENGE

A US financial-services firm wanted to move beyond AI experimentation into production — but their engineers needed concrete, trustworthy patterns and their business leaders needed a shared vocabulary for governance and risk. In financial services, shipping an AI workload without evaluation, guardrails and a clear model-risk story is a non-starter, so speed alone was worthless: the first production workload had to be defensible under internal review, not just functional.

WHAT WE BUILT

- Ran a 3-day intensive: day 1 with engineering on architecture patterns, security and evaluation; days 2-3 building a production-ready workload end-to-end
- Ran parallel sessions with business leaders on AI strategy, governance and prioritisation frameworks so risk ownership was clear
- Delivered a working reference workload with infrastructure-as-code, an evaluation harness and guardrails the team could inspect and trust
- Left an internal playbook the team uses to greenlight and govern subsequent AI initiatives

OUTCOMES

- The team shipped their first production AI workload within the 3 days — with an evaluation harness, guardrails and governance built in from the first commit
- Engineering and business stakeholders aligned on the AI roadmap and its guardrails
- Infrastructure-as-code and evaluation made the workload reproducible and defensible under internal review
- An internal playbook now used to govern and greenlight further AI initiatives

STACK