

The AI-Native Software Benchmark

How mid-market enterprises are building, buying, and operating AI-native systems.

EXECUTIVE SUMMARY

Across our 2025 and 2026 engagements, the gap between organizations experimenting with AI and organizations operating AI in production widened sharply. The leaders are not the ones with the largest model budgets; they are the ones who rebuilt a workflow end to end.

This benchmark summarizes what we observed across delivery work in healthcare, life sciences, financial services, insurance, retail, software, and real estate and home services.

KEY FINDINGS

1. Workflow ownership beats model selection. Teams that assigned a single owner to an end-to-end workflow shipped roughly twice as fast as teams organized around a platform.
2. Data readiness is the true schedule risk. In most programs, the majority of the first phase went to consolidating source systems rather than to model work.
3. Evaluation is the missing discipline. Few teams had a written acceptance bar for AI output before launch, and those without one spent the longest in rework.
4. Human review is a feature, not a fallback. The systems that stayed in production kept a reviewer in the loop on high-consequence steps.

WHAT LEADERS DO DIFFERENTLY

They scope to one measurable process, instrument it before launch, and hold the same reliability bar they would hold for any other production system. They also plan for handover from day one, so the internal team can operate the system without the original build team.

HOW TO USE THIS REPORT

Use the findings as a checklist for your next AI initiative: named workflow owner, documented data sources, written evaluation bar, defined review points, and an explicit handover plan.