Part of the Code Ninety Enterprise AI Research Series 2026 | Published February 14, 2026 | Updated August 2026 | By Babar Khan, Managing Director & Co-Founder, Code Ninety

Enterprise AI PoC Graveyard & Pilot-to-Production Conversion Study 2026

48.6% of enterprise AI pilots are still stuck in proof-of-concept after 6 months, and only 31.4% ever reach production. This original survey of 210 technical leaders identifies exactly what separates the pilots that convert from the ones that die in the PoC graveyard — and the gap comes down almost entirely to four operational factors, not model quality.

Key findings

- Reach production: 31.4% · Stall 6+ months: 48.6% · Abandoned: 20.0%
- Top failure reason: data silos & quality deficits , cited by 62.4%
- Biggest single predictor: C-suite sponsorship — 84.8% of successes had it vs. 31.1% of failures
- Converted pilots decide in 3.8 months avg ; stalled pilots drift for 8.4 months avg
- Best-converting sector: Enterprise Tech & SaaS at 42.3%; worst: Industrial & Energy at 20.0%

Cite this as: Code Ninety. "Enterprise AI PoC Graveyard & Pilot-to-Production Conversion Study 2026." February 2026. codeninety.com/research/poc-graveyard-conversion-study-2026

How was this AI pilot conversion study conducted?

This study surveyed 210 technical leaders (76.4% completion rate from 275 invited) — CTOs/CIOs (36.2%), VPs of Engineering/Heads of Tech (32.4%), Chief Architects (21.0%), and Heads of AI/MLOps Leads (10.4%) — fielded December 1, 2025 through February 5, 2026. Respondent organizations spanned 250-999 employees (22.4%), 1,000-4,999 (41.9%), 5,000-19,999 (23.8%), and 20,000+ (11.9%), across financial services (26.7%), enterprise tech/SaaS (24.8%), healthcare (19.0%), retail (15.2%), and industrial/energy (14.3%), with 51.4% North America, 31.9% Western Europe, and 16.7% Asia-Pacific.

What percentage of AI pilots actually reach production?

Of all AI pilots tracked in this survey, only 31.4% successfully deployed to production. 48.6% remain stalled in the proof-of-concept stage for more than 6 months without a clear go/no-go decision, and 20.0% were abandoned or cancelled outright. Combined, nearly 70% of enterprise AI pilots either die slowly or die outright — production deployment is the minority outcome, not the default one.

Time-to-decision compounds this problem. Pilots that eventually converted to production reached that decision in an average of 3.8 months. Pilots that stalled were, on average, already 8.4 months in without resolution, and abandoned projects ran for 6.1 months before being cancelled. A pilot still undecided past roughly 4 months is trending toward the stalled cohort, not the production cohort — the data shows a clear

inflection point rather than a gradual curve.

What this means: a PoC timeline without a hard decision gate is functionally a slow-motion abandonment, given that stalled pilots average more than double the duration of converted ones. Organizations should treat "still in PoC after 4-5 months" as a decision-forcing signal, not a normal extended evaluation period.

Why do AI pilots actually fail?

Respondents cited data silos and unstructured data quality deficits as the top blocker at 62.4% — ahead of unclear financial ROI or lack of baseline metrics (58.1%), inference latency and production reliability gaps (51.9%), security/governance blockers (46.2%), escalating token and compute cost at scale (39.5%), and lack of dedicated MLOps/AI data engineering talent (35.7%).

What this means: data quality outranking every other blocker, including ROI clarity, is a specific and actionable signal — most pilots aren't failing because the underlying use case is wrong or the model is inadequate, they're failing because the data feeding the model isn't in a usable state. This is consistent with what our own AI project scoping guidance recommends: a data readiness assessment before pilot kickoff, not after, since retrofitting data quality mid-pilot is exactly the kind of unplanned work that pushes a project from the 3.8-month converter timeline into the 8.4-month stalled timeline.

What separates AI pilots that convert from ones that stall?

This is the most useful data in the study: four specific operational factors show a large, measurable gap between the production cohort and the failed cohort.

What this means: pre-defined quantifiable success metrics shows the single widest gap (60.6 points) of any factor measured — more than sponsorship, more than data engineering resourcing. This directly corroborates the "unclear ROI" failure reason cited above (58.1%): pilots without agreed success criteria going in have no objective basis for a go/no-go decision, which is exactly the condition that produces a stalled pilot rather than a decisively killed or converted one. Combined with C-suite sponsorship, these two factors alone are the strongest predictors in the dataset of whether a pilot reaches production.

Which industries convert AI pilots to production most successfully?

Enterprise Tech & SaaS converts pilots to production at the highest rate (42.3%), followed by Financial Services & Banking (33.9%), Retail & Consumer Goods (28.1%), Healthcare & Life Sciences (22.5%), and Industrial & Energy (20.0%).

What this means: this ranking broadly tracks each industry's existing data infrastructure maturity and regulatory complexity — SaaS companies typically already have cloud-native, well-instrumented data pipelines (directly addressing the top failure reason above), while healthcare and industrial sectors carry more legacy system fragmentation and, in healthcare's case, additional regulatory review overhead. An organization in a lower-converting industry shouldn't read this as "AI won't work for us" — it should be read as "the data-readiness and governance work matters more here, budget for it explicitly rather than assuming SaaS-sector pilot timelines will transfer directly."

How does pilot discipline connect to financial ROI?

This study's four production-cohort differentiators — C-suite sponsorship, pre-defined success metrics, dedicated data engineering, and automated guardrails — line up directly with the financial outcomes in our companion Enterprise AI ROI & Payback Timeline Benchmark . That study found only 38.2% of organizations report measurable positive EBIT impact from AI investment overall, while 18.2% report outright negative financial drag. Read alongside this study's finding that 78.8% of production-stage pilots had pre-defined quantifiable success metrics versus just 18.2% of failed pilots, the throughline is direct: a pilot without agreed success criteria isn't just at higher risk of stalling — it's structurally unable to demonstrate the kind of measurable EBIT impact the ROI study shows only a minority of organizations currently achieve, because there's no baseline it was ever measured against.

This reframes "pilot success" and "financial ROI" as the same underlying discipline problem measured at two different stages, not two separate initiatives. An organization investing in the four differentiators this study identifies isn't just improving its odds of reaching production — it's building the measurement infrastructure that the ROI study's positive-EBIT cohort already has in place, which is why treating pilot governance as a pre-launch checkbox rather than an ongoing discipline undersells its financial value.

What's a practical checklist for launching an AI pilot?

Based on the four differentiators measured, an organization launching an AI pilot with production intent should have, before kickoff: named executive sponsorship at the C-suite level (not just a department head); a written, quantifiable success metric agreed before the pilot starts, not retrofitted afterward; dedicated data engineering capacity allocated specifically to the pilot, not borrowed ad hoc from other projects; and automated evaluation/guardrail infrastructure in place before scaling past initial testing. Pilots missing two or more of these four factors should be treated as high risk for the stalled-or-abandoned outcome this data shows affects roughly 69% of pilots overall.

What are this study's methodology and limitations?

This is original primary research from 210 completed survey responses (76.4% completion rate from 275 invited), fielded December 1, 2025 to February 5, 2026 across Code Ninety's client and prospect network of enterprise technical decision-makers.

Limitations: as with our companion AI adoption survey, respondents were drawn from Code Ninety's own network rather than a fully independent random sample, which may skew toward organizations already engaging with external technical partners. The differentiator comparisons (production vs. failed cohort) are correlational, not causal — this data shows a strong association between factors like C-suite sponsorship and production conversion, but doesn't isolate whether sponsorship directly causes better outcomes or is itself a proxy for broader organizational AI readiness. We report this distinction transparently rather than overclaiming causation the data doesn't establish.

Frequently asked questions

What percentage of enterprise AI pilots actually reach production?

Only 31.4% of enterprise AI pilots successfully deploy to production, based on a survey of 210 CTOs, VPs of Engineering, and AI Program Directors. 48.6% stall in the proof-of-concept stage for more than 6 months, and 20.0% are abandoned or cancelled outright.

Why do most AI pilots stall or fail?

Data silos and unstructured data quality deficits are the top-cited failure reason at 62.4%, ahead of unclear financial ROI or lack of baseline metrics (58.1%), inference latency and production reliability gaps (51.9%), and security/governance blockers (46.2%).

What's the single biggest predictor of whether an AI pilot reaches production?

Executive C-suite sponsorship shows the largest gap between successful and failed cohorts: 84.8% of pilots that reached production had C-suite sponsorship, versus just 31.1% of failed pilots — a 53.7 percentage point gap, the widest of any measured differentiator.

How long does it typically take an AI pilot to convert to production?

Pilots that successfully converted to production took an average of 3.8 months to reach that decision. Pilots that stalled averaged 8.4 months still undecided, and abandoned projects averaged 6.1 months before being cancelled — meaning a pilot still undecided past 4 months is trending toward the stalled cohort, not the production cohort.

Which industries convert AI pilots to production most successfully?

Enterprise Tech & SaaS leads at a 42.3% conversion rate, followed by Financial Services & Banking (33.9%), Retail & Consumer Goods (28.1%), Healthcare & Life Sciences (22.5%), and Industrial & Energy (20.0%) — a pattern that broadly tracks each industry's existing data infrastructure maturity.

How was this study conducted?

210 completed responses (76.4% completion rate from 275 invited) from CTOs, VPs of Engineering, Chief Architects, and AI Program Directors across financial services, tech, healthcare, retail, and industrial sectors, fielded December 1, 2025 to February 5, 2026.

Our AI pilot has been running for 5 months with no clear signal — should we kill it now or give it more time?

This data suggests treating 5 months as a decision-forcing point, not a normal extended evaluation window — converted pilots average 3.8 months to decision, while stalled pilots average 8.4 months still undecided. A pilot at 5 months without a clear success/failure signal is already trending toward the stalled cohort's timeline, not the converter's, so the more useful question isn't "kill or continue" but "why don't we have a clear signal yet" — usually a missing pre-defined metric, which this study found in 81.8% of failed pilots.

How do I get executive sponsorship if leadership isn't already bought in?

This study didn't measure how sponsorship was obtained, only its correlation with outcomes, but the size of the gap (84.8% vs. 31.1%) suggests it's worth treating as a prerequisite to secure before significant pilot investment, not something to pursue in parallel with technical work. A pilot framed around a metric a C-suite sponsor already cares about (cost, revenue, risk reduction) is more likely to earn that sponsorship than one framed purely around technical capability.

What does a "good" pre-defined success metric actually look like before starting a pilot?

Based on what separates this study's production cohort from its failed cohort, a good metric is quantifiable before the pilot starts (not retrofitted afterward), tied to a business outcome a sponsor already tracks, and specific enough that both a "yes, this worked" and a "no, this didn't" outcome are equally clear in advance. A metric like "improve customer support efficiency" fails this test; "reduce average ticket resolution time by 20% within 90 days" passes it.

Is a stalled pilot salvageable, or should it just be restarted from scratch?

This study didn't track pilot restarts directly, but given that the top blocker (data silos and quality deficits, 62.4%) and the top differentiator (pre-defined metrics, 78.8% vs 18.2%) are both addressable without discarding existing technical work, a stalled pilot is often salvageable by retrofitting a clear success metric and securing executive sponsorship, rather than requiring a full technical restart. Treat a full restart as necessary only when the underlying use case itself, not just the pilot's governance, was wrong.
