

ML LABS INTELLIGENCE

Which AI Pilots Deserve Production Investment

AI pilots get stuck between "too promising to kill" and "too incomplete to ship." A three-gate graduation framework that moves the right experiments forward and kills the rest.

STRATEGIC

Author Omar Trejo

Date 2026-04-07

ML LABS

mlabs.com/intel/which-ai-pilots-deserve-production

Most organizations have run several AI pilots. Few can say how many are in production delivering measurable value. The pattern is so common it has a name: pilot purgatory. [Research on AI in business](#) (MIT Sloan, 2024) found that most executives believe AI will offer a competitive advantage, yet only a fraction have incorporated it at scale. The gap between belief and deployment is where organizations bleed budget and credibility. This article covers *which* pilots to promote. For the engineering side of getting one into production, see [The Engineering Path From AI Pilot to Production](#).

Why Pilots Get Stuck

Most AI proof-of-concepts succeed at demonstrating feasibility — that's the easy part. **Feasibility and production-readiness are completely different evaluations**, and most organizations conflate them. A pilot that proves "we can predict customer churn with 82% accuracy" has answered a technical question, not an operational one — can the system integrate with the CRM, act on predictions fast enough to matter, and survive data drift after the data science team moves on?

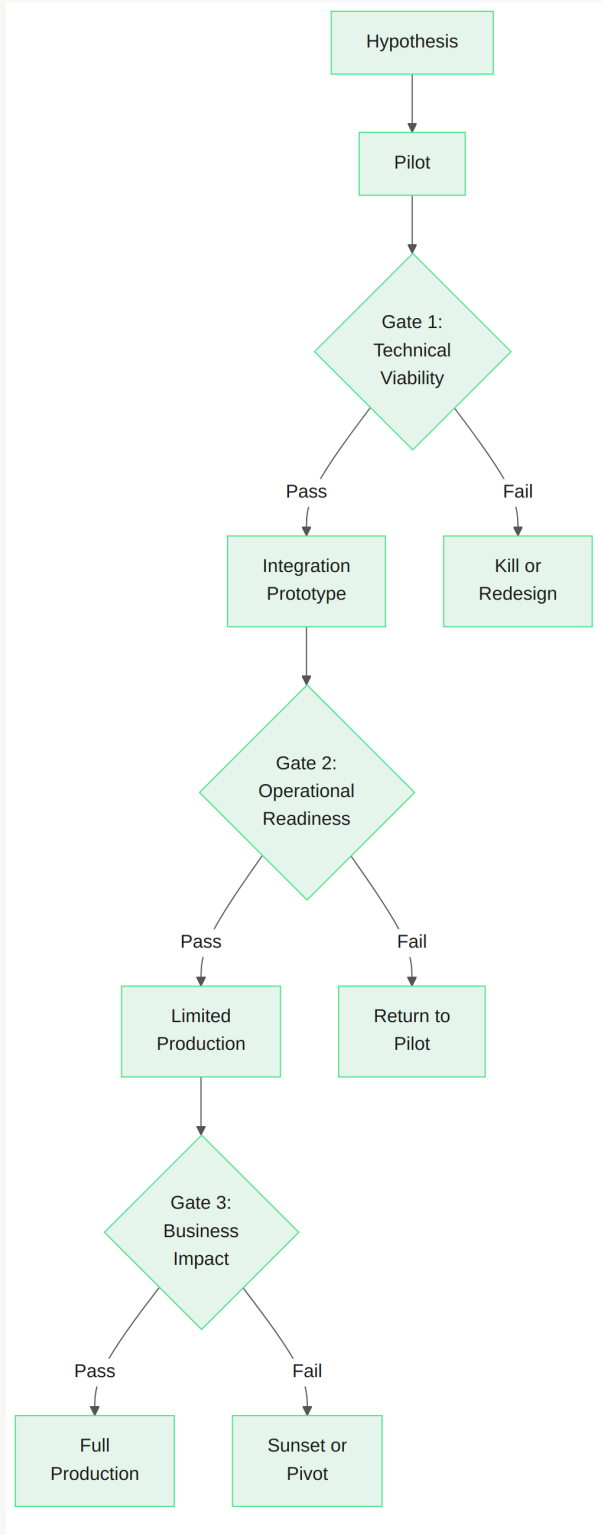
Without clear answers, pilots sit in limbo — too promising to kill, too incomplete to ship.

// The most common failure: a pilot passes all technical checks but has no operational home. The data science team built it and wants to move on. The engineering team didn't build it and doesn't want to maintain it. The system gets deployed with no owner, silently degrades, and eventually fails.

Counterintuitively, larger pilot portfolios often correlate with weaker production outcomes. Concentrated expertise with clear ownership gets systems to production faster than committees with shared responsibility. [Research on AI deployment rates](#) (Deloitte, 2024) confirms that organizations with many concurrent pilots tend to have lower overall deployment rates than those with fewer, more focused ones.

The Graduation Framework

Moving from experimentation to impact requires explicit criteria, clear ownership, and predefined kill conditions. This framework has three gates, each progressively harder to clear. Most organizations only evaluate the first and assume the rest will work itself out. It doesn't.



Gate 1: Technical Viability

Most organizations stop here — a successful demo and the pilot is declared a success. But technical viability is the lowest bar, not the finish line. Verify performance on representative production data (not curated demos), confirm latency meets business process constraints, and ensure another engineer can reproduce the pipeline end-to-end. If representative data performance drops below the business-useful threshold, kill the pilot and document what was learned.

Gate 2: Operational Readiness

Most pilots die here — not from technical failure, but because nobody planned for operations. [Research on ML systems in production](#) (NeurIPS, 2015) showed that surrounding infrastructure constitutes the vast majority of a production system's complexity. The system needs real data sources with defined API contracts, drift monitoring, documented failure procedures owned by a specific team, and a completed security review. If no team will accept production ownership, return to pilot or kill. This gate is genuinely hard the first time; pattern recognition from prior deployments compresses the learning curve significantly.

Gate 3: Business Impact

Evaluated after a defined production window — long enough to accumulate signal, short enough to stay accountable. Measure key business metrics against pre-deployment baselines, confirm users are actually using the system with positive cost-benefit, and verify the system handles peak load. If business impact can't be demonstrated within the window, sunset the system — sunk costs are irrelevant.

The Zombie Problem

Between Gate 2 and Gate 3 lies the most dangerous zone: systems that are technically deployed but not generating value. Zombies are more expensive than failed pilots because they incur ongoing costs indefinitely — every deployed system needs a mandatory impact review with a short remediation window before decommissioning.

Post-Graduation

Passing Gate 3 is the beginning of the next phase. **Recent data on AI investment trends** (Stanford HAI, 2025) underscores that scaling validated systems is where the real value lies. Evaluate expansion potential, infrastructure readiness for full load, and whether ROI strengthens at volume. Successful graduates build organizational confidence that improves future pilot selection — the first graduation is always the hardest.

Making the Framework Stick

Research on scaling AI (McKinsey, 2024) found that organizations with dedicated AI leadership are significantly more likely to achieve production deployments. The **AI Risk Management Framework** (NIST, 2023) provides a complementary governance structure. Invest proportionally to gate progress — minimal pre-Gate 1, moderate between Gates 1-2, full between Gates 2-3. Three structural changes sustain the discipline: appoint a graduation owner with kill authority at each gate (not a committee), cap concurrent pilots at 3-5, and make killing easy and blame-free.

Expected Results

Research on structured AI scaling frameworks (HBR, 2025) found that systematic approaches reduced project delivery times by an estimated 50-60%. Resources concentrate on the most promising candidates, clear criteria eliminate ambiguity that causes stalls, and explicit kill criteria prevent zombie projects from accumulating.

When This Approach Does Not Apply

This framework loses its power when leadership insists every pilot must continue regardless of results — gate reviews happen on schedule, but no pilot is ever killed. A related failure occurs when the graduation owner lacks real authority and department heads can override kill recommendations. Organizations in this position need to resolve the authority question first: quantify the cost of zombie systems, calculate resources trapped in non-producing pilots, and present the opportunity cost in terms leadership can act on.

First Steps

1. **Inventory active pilots.** Apply Gate 1 retroactively. If you cannot enumerate them in 10 minutes, that is your first problem. Viability checks alone typically eliminate a significant portion.
2. **Find the zombies.** Check Gate 3 criteria on anything already deployed. If you cannot demonstrate measurable business impact with data, the system is a zombie candidate.

3. **Assign a graduation owner.** Give one person the authority to evaluate pilots against gate criteria on a 90-day cycle, with standing to kill projects without career risk.

Practical Solution Pattern

Adopt explicit graduation gates so pilots are either promoted with evidence or shut down with documented learning. Assign one person the authority to approve or kill at each gate, cap concurrent pilots at three to five, and enforce a mandatory 90-day business impact review for every deployed system.

This works because it eliminates the two conditions that sustain pilot purgatory: ambiguous promotion criteria and diffuse accountability. When graduation criteria are defined in advance, teams know exactly what "done" means. When kill decisions are blame-free and routinized, the portfolio stays lean and organizational capacity for production work remains intact. If you need to decide which experiments deserve production budget, a **Strategic Scoping Session** can turn that portfolio question into a written recommendation and next step.

References

1. MIT Sloan Management Review. **Artificial Intelligence in Business Gets Real.** *MIT Sloan Management Review*, 2024.
2. Deloitte. **State of AI in the Enterprise.** *Deloitte Insights*, 2024.
3. Sculley, D., et al. **Hidden Technical Debt in Machine Learning Systems.** *NeurIPS*, 2015.
4. Stanford HAI. **2025 AI Index Report.** *Stanford Human-Centered Artificial Intelligence*, 2025.

5. McKinsey & Company. **The State of AI.** *McKinsey Global Survey*, 2024.
 6. NIST. **AI Risk Management Framework.** *National Institute of Standards and Technology*, 2023.
 7. Harvard Business Review. **Most AI Initiatives Fail. This 5-Part Framework Can Help.** *Harvard Business Review*, 2025.
-



ML LABS

Custom AI Systems for High-Value Workflows

mllabs.com