



White Paper

From Pilots to Production

The POC worked. Now what?
Why scaling AI from a sandbox to the real organization is a people problem, not a technology problem.

Alexander S. Petty

SINGULARICS | singularics.ai

© 2026 SINGULARICS. All rights reserved.

No part of this publication may be reproduced, distributed, or transmitted in any form or by any means without the prior written permission of SINGULARICS.

For permissions, bulk copies, or licensing inquiries:

hello@singularics.ai

First published April 2026.

singularics.ai

Contents

1	The Standing Ovation That Changed Nothing	4
2	The Sandbox Illusion	5
3	The Five Walls	7
3.1	Wall 1: Data readiness	7
3.2	Wall 2: Integration complexity	8
3.3	Wall 3: Governance approval	8
3.4	Wall 4: Organizational ownership	8
3.5	Wall 5: Behavioral adoption	9
4	The Sandbox Teaches Almost Nothing	10
5	The POC Addiction	11
6	Designing for Production from Day One	12
6.1	Start with real data	12
6.2	Involve the operating team from day one	12
6.3	Address governance early	12
6.4	Test with skeptics, not enthusiasts	12
6.5	Measure behavioral change, not model performance	13
7	The Real Measure of Success	13

Abstract

The enterprise has a POC addiction. Build a proof of concept. Demo it for the steering committee. Get applause. Declare success. Move on to the next POC. Meanwhile, nothing in production has changed. This paper examines the five structural walls between a successful pilot and production deployment, why organizations systematically underestimate them, and what it takes to cross the gap that separates a compelling demo from organizational value.

The Standing Ovation That Changed Nothing

The demo was flawless.

A product team at a mid-cap insurance company had spent eight weeks building an AI-powered claims triage system. The pilot processed 200 sample claims with 92% accuracy. It reduced initial assessment time from 45 minutes to 4 minutes. The projected annual savings, extrapolated across the full claims volume, was \$14 million.

The steering committee was impressed. The CTO called it “the most promising AI initiative we’ve seen.” The pilot team received recognition at the quarterly all-hands. Budget was approved for “scaling to production.”

Fourteen months later, the system was not in production. The pilot team had been reassigned. The projected \$14 million in savings existed only in the original presentation. The claims department was processing every claim exactly the way it had before the demo.

What happened?

Nothing dramatic. No catastrophic failure. No executive reversal. The system simply encountered the reality that the sandbox had been carefully designed to avoid: messy data, complex integrations, compliance requirements, organizational ownership questions, and the behavioral change required for 300 claims adjusters to work differently.

The pilot succeeded. The organization did not change around it.

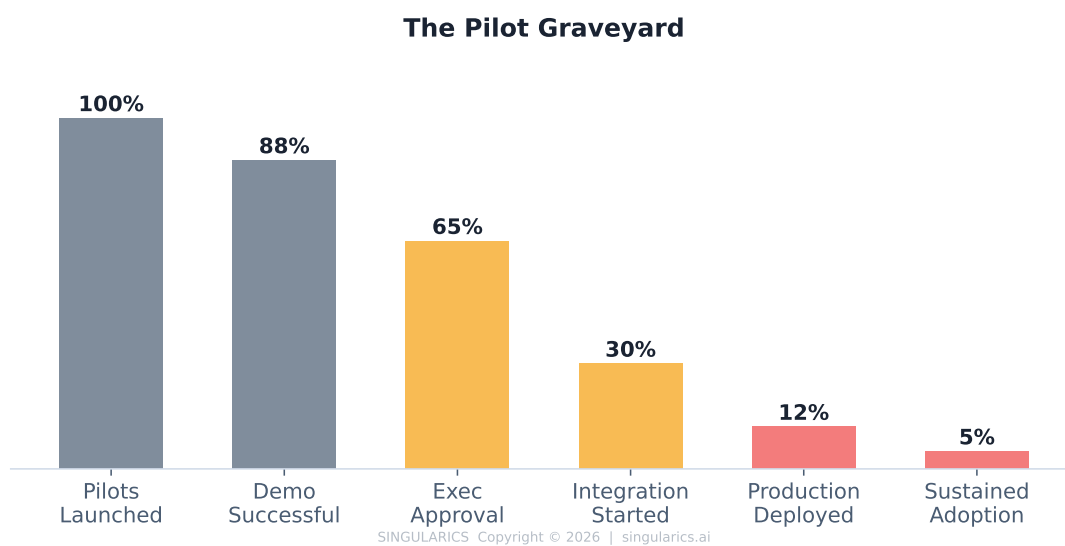


Figure 1: The pilot graveyard: of 100 AI pilots launched, only 5 result in sustained production adoption. The attrition occurs not at the technology layer but at the organizational one.

95% of generative AI pilots at companies do not reach production deployment. The technology works in the sandbox. The organization cannot absorb it. [1]

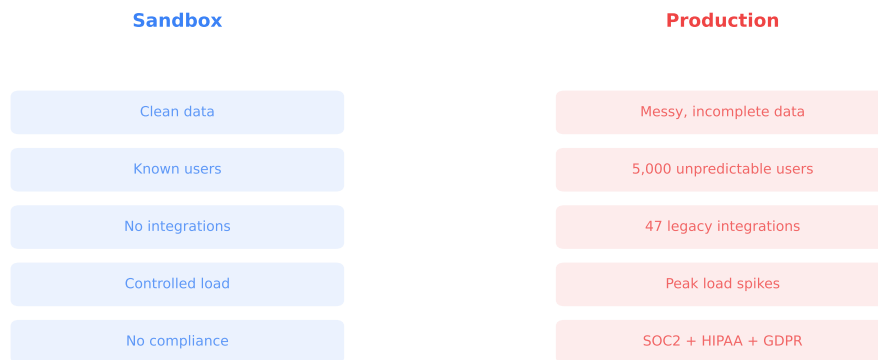
The Sandbox Illusion

A proof of concept is designed to answer one question: *can this technology do what we need it to do?*

This is a valid question. It is also the wrong question to answer first.

The more important question is: *can our organization absorb this technology into how it actually operates?*

A POC cannot answer this question because a POC is designed to eliminate exactly the variables that determine the answer. The data is clean. The users are selected. The integrations are mocked. The compliance requirements are deferred. The organizational change is out of scope.



SINGULARICS Copyright © 2026 | singularics.ai

Figure 2: The conditions that make a pilot succeed are precisely the conditions that production will not provide. Every variable that the sandbox eliminates is a variable that production will impose.

The sandbox is a controlled experiment. Production is an uncontrolled environment. The distance between them is not a gap to be crossed. It is a transformation to be managed. And most organizations have not scoped, staffed, or funded that transformation.

A successful pilot proves that the technology works under conditions you will never encounter again. Production is where you encounter every condition the pilot was designed to avoid.

This is why the steering committee's enthusiasm after a pilot demo is one of the most dangerous moments in an AI initiative. The enthusiasm creates a false sense of proximity to value. "It works! Now we just need to scale it." The word "just" in that sentence conceals months of work that nobody has estimated, budgeted, or assigned.

The Five Walls

Between a successful pilot and sustained production adoption, there are five walls. Each one is capable of stopping an initiative entirely. Most organizations hit at least three.

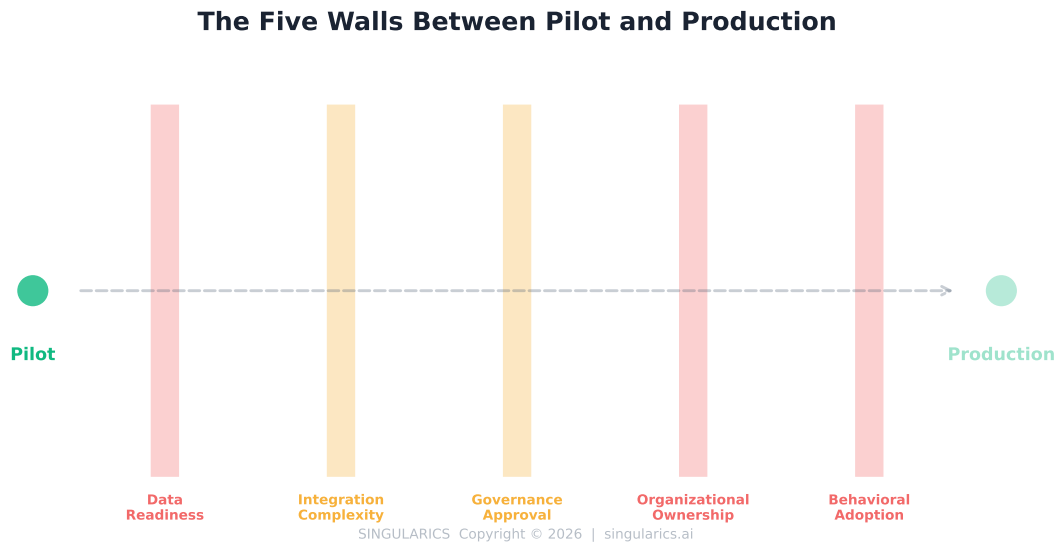


Figure 3: Five structural barriers stand between a successful pilot and production. The technology clears the first wall easily. It is the remaining four that determine whether the organization captures value.

Wall 1: Data readiness

The pilot used curated data. Production requires the data as it actually exists: incomplete, inconsistent, spread across systems that were never designed to talk to each other.

The claims triage pilot used 200 carefully selected claims with complete documentation. The production claims database contained 2.3 million records, 40% of which had missing fields, inconsistent formatting, or data quality issues that the pilot never encountered.

Data readiness is not a data engineering problem alone. It is a governance problem, a process problem, and often a political problem. The data the AI needs may live in a system owned by a department that was not involved in the pilot, has different priorities, and has no incentive to cooperate.

Wall 2: Integration complexity

The pilot operated in isolation. Production requires integration with the systems, workflows, and processes that already exist.

The claims triage system needed to integrate with the claims management platform, the document management system, the customer communication workflow, the compliance audit trail, and the existing escalation process. Each integration introduced technical complexity, organizational negotiation, and timeline risk.

Integration is where pilot timelines die. The pilot took eight weeks. The integration assessment alone took twelve.

Wall 3: Governance approval

The pilot operated under a research exemption. Production requires full governance approval.

This means security review, data privacy assessment, compliance certification, risk classification, and often regulatory approval. Each of these processes has its own timeline, its own stakeholders, and its own criteria. None of them were engaged during the pilot because the pilot was explicitly scoped to avoid them.

As explored in our research report on governance, the approval process itself can become the barrier. When governance is designed as a gate rather than a guardrail, the time from “pilot success” to “governance approval” can exceed the time it took to build the pilot in the first place.

Wall 4: Organizational ownership

The pilot was owned by an innovation team. Production requires ownership by the business unit that will operate it.

This transfer of ownership is one of the most underestimated challenges in the pilot-to-production journey. The innovation team built the system. They understand it. They are invested in it. The business unit was not involved in building it. They did not choose it. They may not want it.

Ownership transfer requires the business unit to accept responsibility for a system they did not design, staff it with people who did not build it, and maintain it with a budget they did not plan for. This is not a handoff. It is a persuasion campaign.

The innovation team builds the pilot. The business unit must operate the production system. These are different organizations with different priorities, different incentives, and different definitions of success. The handoff between them is where most pilots go to die.

Wall 5: Behavioral adoption

The pilot was used by enthusiasts. Production requires adoption by everyone, including the people who did not ask for it, do not understand it, and may feel threatened by it.

Three hundred claims adjusters did not participate in the pilot. They did not experience the excitement of building something new. They experienced a memo telling them their workflow was changing. Some embraced it. Some ignored it. Some actively resisted it.

This is the wall where the organizational change management that was never funded, never staffed, and never scoped finally becomes visible. The technology works. The people have not been prepared to use it. And preparing 300 people to change how they work is a fundamentally different activity than building a system that processes claims.

The Sandbox Teaches Almost Nothing

Organizations invest heavily in pilots because pilots feel like learning. The team is experimenting. Data is being gathered. Capabilities are being assessed. It feels productive.

But the learning that a sandbox produces is almost entirely about the technology. It answers: can the model do this task? How accurate is it? What are its failure modes under controlled conditions?

These are useful answers. They are also the easy answers. The hard answers, the ones that determine whether value is captured, can only be learned in production.

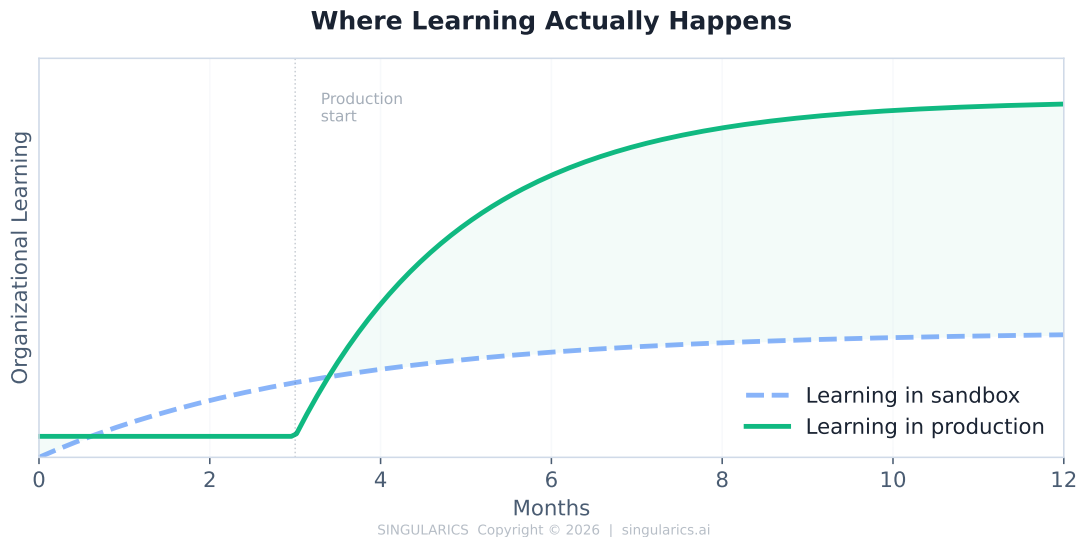


Figure 4: Organizational learning in a sandbox plateaus quickly. The learning that compounds, the learning about how work changes, how people adapt, how systems integrate, only begins when AI enters production.

Production teaches you that the model behaves differently with messy data. That users interact with it in ways the designers did not anticipate. That the workflow you designed does not match the workflow people actually follow. That the edge cases you dismissed as unlikely happen every day. That the integration you assumed would be straightforward requires renegotiating a contract with a vendor.

None of this learning is available in the sandbox. And it is precisely this learning that determines whether the initiative creates value.

You will learn more in the first week of production than in three months of piloting. Not because production is better designed. Because production is real, and reality is the only teacher that does not grade on a curve.

The POC Addiction

If pilots teach so little about production readiness, why do organizations keep running them?

Three reasons.

Pilots are safe. A pilot cannot fail in a way that harms the organization. The data is sandboxed. The users are selected. The blast radius is contained. This safety is the pilot's value and its limitation.

Pilots are fundable. A pilot requires a small budget, a small team, and a short timeline. It fits within discretionary spending. It does not require board approval. It does not require organizational change. It is the path of least resistance.

Pilots produce demos. A demo is a powerful social artifact. It creates excitement. It generates executive attention. It produces the feeling of progress. The demo is the reward, and the reward cycle, from idea to demo in eight weeks, is intoxicating.

The result is an organization that runs pilot after pilot, produces demo after demo, and never crosses the gap to production. The innovation portfolio grows. The production portfolio does not. The organization accumulates proof that AI *could* work without ever proving that it *does* work.

46% of companies abandoned at least one AI initiative in 2025, with an average sunk cost of \$7.2 million per abandoned initiative. Most were technically successful pilots that never reached production. [2]

Designing for Production from Day One

The alternative to the pilot-then-scale model is to design for production from the beginning.

This does not mean skipping experimentation. It means structuring experimentation so that the learnings are transferable to production and the barriers to production are addressed during the experiment, not after.

Start with real data

Use production data from day one. Not a curated subset. Not a synthetic dataset. The actual data, with its actual quality issues, its actual access constraints, and its actual governance requirements.

This is harder than using clean data. It is also more honest. The data quality issues that will block production are discoverable in week one instead of month eight.

Involve the operating team from day one

The business unit that will own the production system must be involved in building it. Not as reviewers at the end. As co-builders from the beginning.

When the operating team helps build the system, they understand it. They have ownership. They have investment. The handoff is not a handoff. It is a continuation of work they have been doing all along.

Address governance early

Engage the governance process during the experiment, not after. Identify the data classifications, the compliance requirements, the security constraints, and the regulatory considerations while there is still time to design around them.

The worst outcome is a technically brilliant pilot that governance cannot approve. Early engagement eliminates this scenario.

Test with skeptics, not enthusiasts

Pilots staffed with volunteers and enthusiasts produce optimistic results. Production will be operated by everyone, including the reluctant, the skeptical, and the resistant.

Include skeptics in the experiment. Their feedback is more valuable than the enthusiasts' feedback, because it reveals the barriers that production will impose.

An enthusiast tells you what is possible. A skeptic tells you what will actually happen.

Measure behavioral change, not model performance

Model accuracy in the sandbox is a necessary but insufficient condition for production value. The real metric is whether the people using the system are working differently and producing better outcomes.

Design the experiment to measure behavior from the beginning. Are decisions being made faster? Is output quality changing? Are workflows actually operating as redesigned? These are the metrics that predict production success. Model performance metrics do not.

The Real Measure of Success

A pilot is successful when the demo goes well. Production is successful when work has changed.

These are different definitions of success, and they require different measurement. The organization that celebrates pilot demos without tracking production adoption is celebrating the appetizer and skipping the meal.

The real measure of an AI initiative is not whether it can work. It is whether it does work, every day, operated by the people who were not in the room when the demo happened, using the data that was not in the sandbox, under the conditions that the pilot was designed to avoid.

That is production. That is where value lives. Everything before it is rehearsal.

The demo is the appetizer. Production is the meal. Most organizations have a portfolio full of appetizers and an empty table.

Pilots succeeding but nothing reaching production? We can help.

30 minutes to talk about crossing the gap. No pitch.

[Book a call](#) | singularics.ai

SINGULARICS

We help large organizations close the intent gap between strategy and execution so that when they accelerate with AI, they accelerate in the right direction.

singularics.ai | [Book a conversation](#)

References

- [1] MIT Sloan Management Review (2025). *Scaling Generative AI: From Pilot to Production*. MIT Sloan Research Report.
- [2] Deloitte (2025). *State of AI in the Enterprise*, 8th Edition. Deloitte Insights.
- [3] Rogers, E.M. (2003). *Diffusion of Innovations*, 5th Edition. Free Press.
- [4] Christensen, C.M. (1997). *The Innovator's Dilemma*. Harvard Business Review Press.
- [5] Gartner (2026). *Hype Cycle for Artificial Intelligence, 2026*. Gartner Research Report.