



It's Almost Never the Model

Why most enterprise generative AI shows no P&L impact, and the discipline behind the few that do

Contents

1. The question	3
2. What the headline numbers measure	3
3. Where the value is lost	5
4. What the successful minority does differently	5
5. A discipline for the next project	6
Appendix A: Open Questions	7
Appendix B: Source Notes	8
References	8

1. The question

Over the past three years, companies have spent heavily on generative AI, the class of models that produce text, code and images from a written prompt. Surveys from the same period report that most of this spending has not yet changed the bottom line. This paper asks why that is, and what the minority that does report returns is doing differently.

The evidence has a limit the reader should know at the outset. Nearly all of it comes from surveys of executives run by consulting firms, analyst firms and one university research group. The figures are self-reported, only RAND's report among them is peer reviewed, and each survey defines value in its own way. This paper therefore relies on the direction the studies share rather than on any single number.

That direction is consistent. Where generative AI fails to pay off, the cause is rarely the model itself. It is the choice of problem, the data and systems around the model, and above all whether anyone changed the work the model was meant to support.

2. What the headline numbers measure

The most quoted figure comes from Project NANDA at MIT. Its July 2025 report stated that "95% of organizations are getting zero return" on an estimated \$30 to \$40 billion of enterprise investment ^[1]. The report rests on a review of more than 300 public AI initiatives, structured interviews with representatives of 52 organizations and survey responses from 153 senior leaders, gathered between January and June 2025 ^[1].

The report's own adoption funnel tells a narrower story. For general-purpose tools such as chat assistants, 80 percent of organizations investigated them, 50 percent piloted them and 40 percent put them into use ^[1]. For embedded, task-specific tools, the figures were 60, 20 and 5 percent. Success there meant tools that users or executives described as causing "a marked and sustained productivity and/or P&L impact" ^[1].

Read against that funnel, the 95 percent counts every organization that never ran a task-specific pilot, as well as those whose pilots failed. Of the organizations that did run a pilot, roughly one in four succeeded under a demanding definition of success ^[2]. Critics have also pointed to the small sample, the lack of peer review and the way the headline figure was derived ^{[2][3]}. The report is better read as evidence that custom AI rarely reaches production than as a precise failure rate.

Other studies measure value differently and arrive in the same territory. Table 1 sets them side by side, with what each one actually asked.

Study	What it measured	Finding
McKinsey, 1,491 respondents, July 2024 ^[4]	Tangible impact of generative AI on enterprise-level EBIT	More than 80 percent saw none
BCG, 1,000 executives in 59 countries, 2024 ^[5]	Tangible value from AI	74 percent had yet to show it; 26 percent had the capabilities to move beyond proofs of concept
IBM, 2,000 chief executives, February to April 2025 ^[6]	Return on AI initiatives over recent years	25 percent delivered the expected return; 16 percent scaled across the enterprise
S&P Global Market Intelligence, more than 1,000 respondents in North America and Europe, 2025 ^[7]	Abandonment of AI initiatives	42 percent abandoned most of them, up from 17 percent a year earlier; the average organization scrapped 46 percent of its proofs of concept
Bain, 197 respondents, third quarter of 2025 ^[8]	Expectations met, and links to revenue or cost	80 percent of use cases met or exceeded expectations; 23 percent of respondents could tie initiatives to new revenue or lower costs
McKinsey, 1,719 respondents, May and June 2026 ^[9]	Any EBIT impact attributed to AI	37 percent attributed some; 6 percent were high performers, with 5 percent of EBIT or more and significant value

Table 1. Six surveys of enterprise AI value, with the question each one asked. Source: the studies cited in the first column.

Two forecasts point the same way. In July 2024, Gartner predicted that at least 30 percent of generative AI projects would be abandoned after proof of concept by the end of 2025 ^[10]. In June 2025, it predicted that over 40 percent of agentic AI projects would be canceled by the end of 2027 ^[11]. Agentic AI refers to systems that plan and carry out multi-step tasks with some autonomy.

The Bain result explains part of the spread between these numbers. Most use cases met or exceeded respondents' expectations, yet fewer than a quarter of respondents could connect their initiatives to revenue or cost ^[8]. Satisfaction with a tool and impact on the profit and loss statement are different measurements. Much of the apparent contradiction between surveys comes from mixing the two.

3. Where the value is lost

BCG asked companies where their implementation challenges came from. Around 70 percent stemmed from people- and process-related issues, 20 percent from technology and only 10 percent from the AI algorithms themselves ^[5]. On this account, the model is the smallest part of what goes wrong.

A 2024 RAND study reached a similar conclusion from the builders' side. Its researchers interviewed 65 experienced data scientists and engineers and identified five root causes of failure ^[12]. Stakeholders misunderstand or miscommunicate the problem, or set the wrong measure of success. The organization lacks the data. Technical teams chase new technology instead of the problem. Infrastructure for data and deployment is underfunded. Finally, some problems exceed what the technology can do, and only this last cause concerns the technology's limits.

Gartner's explanations for abandonment run along the same lines. For generative AI it named poor data quality, inadequate risk controls, escalating costs and unclear business value ^[10]. For agentic AI it named escalating costs, unclear business value and inadequate risk controls ^[11]. IBM's survey of chief executives adds a structural cause: half of them said the pace of recent investment had left their organizations with "disconnected, piecemeal technology" ^[6].

The MIT report describes the core barrier as a learning gap. It found that "most GenAI systems do not retain feedback, adapt to context, or improve over time" ^[1]. Users accept such tools for simple tasks and turn away from them for work that matters. The same report found that only 40 percent of companies had purchased an official subscription to a large language model, while workers at over 90 percent of the companies surveyed used personal AI tools for work ^[1]. Employees were finding value in general tools while the official programs struggled to fit the work.

Bain's respondents described the same pattern from the inside. About a third of those who were disappointed said the technology had worked at the pilot level and then failed to scale ^[8].

4. What the successful minority does differently

McKinsey tested 25 organizational attributes against reported impact on EBIT. It found that "the redesign of workflows has the biggest effect" on whether an organization sees EBIT impact from generative AI ^[4]. In its 2026 survey, nearly three-quarters of high performers reported fundamentally redesigning workflows because of AI, against one-quarter of other respondents ^[9].

Measurement separates the two groups as well. Of 12 adoption and scaling practices in McKinsey's 2024 survey, tracking well-defined key performance indicators for generative AI had the most impact on the bottom line, yet fewer than one in five respondents said their organizations did it ^[4]. In 2026, high performers were twice as likely as others to report defined processes for measuring the impact of their AI initiatives ^[9].

BCG's leaders allocate their effort in line with where the problems lie. They put 10 percent of their resources into algorithms, 20 percent into technology and data, and 70 percent into people and processes ^[5]. They also locate most of their value in core business functions, 62 percent, against 38 percent in support functions ^[5].

The MIT data favor partnership over building alone. In its sample, external partnerships with customized tools that learn from use reached deployment about 67 percent of the time, against about 33 percent for internally built tools ^[1]. Asked how they would allocate a budget, executives favored sales and marketing, while some of the largest cost savings the report documented came from automating back-office work ^[1].

RAND recommends technical staff who understand the business purpose, at least a year of commitment to a problem, investment in infrastructure, and realistic expectations ^[12].

5. A discipline for the next project

Taken together, these are management decisions, not technology choices. Skipping them, more than picking the wrong model, appears to separate the majority from the minority.

1. **Start from a line in the profit and loss statement and measure it before the pilot.** A pilot without a baseline can satisfy its users and still prove nothing about value ^{[4][8]}.
2. **Redesign the workflow before scaling the tool.** Workflow redesign is the attribute most strongly associated with EBIT impact in McKinsey's data ^{[4][9]}.
3. **Budget for people and process first.** BCG's leaders do, and that is also where most implementation problems arise ^[5].
4. **Settle data, integration and risk controls before the pilot.** All three recur across these lists of causes ^{[6][10][12]}.
5. **Decide in stages, with criteria for continuing or stopping agreed in advance.** Many proofs of concept will stop, as S&P's figures show ^[7], and escalating costs are a leading reason agentic projects are canceled ^[11]. Agreed criteria let a team stop early and on purpose.
6. **Choose between building and buying deliberately.** In MIT's sample, internally built tools reached deployment about half as often as partnered ones ^[1].

Appendix A: Open Questions

How much of the gap is measurement? The studies define value as EBIT impact, expected return, satisfaction or deployment. None of these figures is audited, and the share of failure that reflects measurement rather than performance is unknown.

Will agentic systems change the pattern? Gartner forecasts widespread cancellations of agentic projects ^[11]. McKinsey reports that the share of respondents from large organizations whose companies scale AI agents rose from 27 percent to 40 percent in a year ^[9]. Whether that scaling produces EBIT impact is not yet measured.

Are the observation windows long enough? The MIT data cover six months ^[1], while RAND recommends committing to a problem for at least a year ^[12]. Some projects counted as failures may simply have been measured early.

Appendix B: Source Notes

Section 2, the MIT figure. The report's executive summary states that 95 percent of organizations are getting zero return. Its own funnel for task-specific tools (60 percent investigated, 20 percent piloted, 5 percent successful) implies that about one in four organizations that ran a pilot succeeded ^{[1][2]}. The prose reports both and uses the funnel for any statement about pilots.

Section 2, the two McKinsey surveys. The 2024 survey asked about tangible impact of generative AI on enterprise-level EBIT, and more than 80 percent reported none ^[4]. The 2026 survey asked whether respondents attribute at least some EBIT impact to AI of any kind, and 37 percent did ^[9]. The questions and samples differ, so the two figures are not a trend.

References

1. Aditya Challapally, Chris Pease, Ramesh Raskar, Pradyumna Chari. "The GenAI Divide: State of AI in Business 2025." MIT NANDA, July 2025. <https://webberwentzel.com/News/Documents/2025/MIT1757411281972.pdf>. Accessed October 5, 2026.
2. Robert Wiblin. "The story behind the bad AI stat that moved markets and misled millions (podcast episode)." 80,000 Hours, April 28, 2026. <https://80000hours.org/podcast/episodes/ai-workplace-mit-study/>. Accessed October 5, 2026.
3. Ali Azhar. "MIT Report Flags 95% GenAI Failure Rate, But Critics Say It Oversimplifies." BigDATAwire (HPCwire), September 4, 2025. <https://www.hpcwire.com/bigdatawire/2025/09/04/mit-report-flags-95-genai-failure-rate-but-critics-say-it-oversimplifies/>. Accessed October 5, 2026.
4. Alex Singla, Alexander Sukharevsky, Lareina Yee, Michael Chui, Bryce Hall. "The state of AI: How organizations are rewiring to capture value." McKinsey & Company, March 12, 2025. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-how-organizations-are-rewiring-to-capture-value>. Accessed October 5, 2026.
5. "AI Adoption in 2024: 74% of Companies Struggle to Achieve and Scale Value." Boston Consulting Group, October 24, 2024. <https://www.bcg.com/press/24october2024-ai-adoption-in-2024-74-of-companies-struggle-to-achieve-and-scale-value>. Accessed October 5, 2026.
6. "IBM Study: CEOs Double Down on AI While Navigating Enterprise Hurdles." IBM Newsroom (IBM Institute for Business Value study), May 6, 2025. <https://newsroom.ibm.com/2025-05-06-ibm-study-ceos-double-down-on-ai-while-navigating-enterprise-hurdles>. Accessed October 5, 2026.



7. Lindsey Wilkinson. "AI project failure rates are on the rise: report." CIO Dive, March 14, 2025. <https://www.ciodive.com/news/AI-project-fail-data-SPGlobal/742590/>. Accessed October 5, 2026.
8. Gene Rapoport, Sanjin Bicanic, Muiyiwa Talabi. "Executive Survey: AI Moves from Pilots to Production." Bain & Company, November 24, 2025. <https://www.bain.com/insights/executive-survey-ai-moves-from-pilots-to-production/>. Accessed October 5, 2026.
9. Dan Tinkoff, Lieven Van der Veken, Michael Chui, Tara Balakrishnan. "The state of AI in 2026: On the road to ROI." McKinsey & Company, August 25, 2026. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>. Accessed October 5, 2026.
10. "Gartner Predicts 30% of Generative AI Projects Will Be Abandoned After Proof of Concept By End of 2025." Gartner, July 29, 2024. <https://www.gartner.com/en/newsroom/press-releases/2024-07-29-gartner-predicts-30-percent-of-generative-ai-projects-will-be-abandoned-after-proof-of-concept-by-end-of-2025>. Accessed October 5, 2026.
11. "Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027." Gartner, June 25, 2025. <https://www.gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027>. Accessed October 5, 2026.
12. James Ryseff, Brandon F. De Bruhl, Sydne J. Newberry. "The Root Causes of Failure for Artificial Intelligence Projects and How They Can Succeed: Avoiding the Anti-Patterns of AI." RAND Corporation, August 13, 2024. https://www.rand.org/pubs/research_reports/RRA2680-1.html. Accessed October 5, 2026.



PRESENTED BY

Prepared with AI assistance; edited by Stefan Brunner.



Stefan Brunner — angel investor, operator, and founder of Westlake Professionals, Austin.

AI in production. Not in pilots. · westlakeprofessionals.com ·

hello@westlakeprofessionals.com

It's Almost Never the Model · Insights · Enterprise AI · Westlake Professionals · October 5, 2026 · Tags: enterprise AI, generative AI, ROI, workflow redesign · Version 1.0