Enterprise AI Deployment Failures and Outcomes in 2026

IntuitionLabs.ai · Adrien Laurent · 2026-09-05 · enterprise ai · ai deployment failures · ai roi · ai pilot to production · ai project failure rate · ai adoption metrics · ai implementation taxonomy · life sciences ai · ai maturity model

Executive Summary

Enterprise investment in artificial intelligence (AI) has outpaced enterprises' ability to convert that investment into measured financial return, and the published failure-rate statistics vary widely because each study defines "pilot," "deployment," and "success" differently. The Massachusetts Institute of Technology (MIT) NANDA initiative's 2025 review of over 300 public AI initiatives, 52 structured interviews, and 153 conference survey responses found that 95% of organizations piloting generative AI report zero measurable profit-and-loss return (nanda.media.mit.edu). The RAND Corporation's own 65-interview study, fielded between August and December 2023, found that 84% of practitioners cited leadership-related causes, not model performance, as the primary reason AI projects fail (rand.org). S&P Global Market Intelligence's survey of 1,006 IT and business professionals found that AI-initiative abandonment before production rose from 17% in 2024 to 42% in 2025, with an average of 46% of projects scrapped between proof of concept and broad adoption (spglobal.com). Gartner separately predicted that at least

30% of generative AI projects would be abandoned after proof of concept by the end of 2025, citing poor data quality, inadequate risk controls, escalating cost, and unclear business value (gartner.com), and later issued a similar prediction for agentic AI specifically, detailed in Table 1 below.

Financial-outcome data tell a related but distinct story from failure-rate data. McKinsey's November 2025 global survey (n=1,993) found that only 39% of respondents attribute any level of earnings-before-interest-and-taxes (EBIT) impact to AI enterprise-wide, with roughly 6% qualifying as "high performers" reporting 5% or more EBIT impact (mckinsey.com). IBM's Institute for Business Value found an average enterprise AI return on investment (ROI) of just 5.9%, below the roughly 10% cost of capital that most firms require, with fewer than one in four organizations exceeding 10% ROI (ibm.com). Deloitte's October 2025 survey of 1,854 executives found only about one in five organizations qualify as "AI ROI Leaders" (deloitte.com), and BCG's September 2025 study found that only 5% of companies are "future-built" for AI, capturing 1.7 times the revenue growth and 3.6 times the three-year total shareholder return of the 60% of companies it classifies as laggards (bcg.com).

This report builds a source-linked evidence base rather than a single headline statistic. It defines the terms (pilot, deployment, adoption, quality, and financial outcome) that make these studies comparable, states the method and assumptions behind each figure it cites, and compares the studies side by side in reproducible tables rather than repeating any single number as if it were the whole picture. It also documents the pilot-to-production gap directly: an IDC survey of 900 organizations in Asia-Pacific, commissioned by Lenovo, found fewer than 10% of proof-of-concept projects were deemed successful against predefined business goals (lenovo.com), and O'Reilly's November 2023 survey found that while most organizations were already using generative AI, only a small minority had reached production, detailed in the Analysis of Key Segments section below.

Life sciences offers a narrower but instructive segment of this picture. Sanofi's 2023 announcement of a company-wide AI push described a target of becoming "the first pharma company powered by artificial intelligence at scale," reporting research-process timelines cut from weeks to hours in some cases and an inventory-forecasting tool it said could predict 80% of low-inventory positions (sanofi.com). A peer-reviewed survey of 67 non-profit US health systems, fielded in fall 2024, found that immature AI tooling was the most commonly cited barrier to deployment even as every responding system had begun piloting ambient clinical-documentation AI (pmc.ncbi.nlm.nih.gov). Read together, the evidence supports a specific, bounded conclusion: enterprise AI failure is a real and well-documented phenomenon at the proof-of-concept and value-realization stages, its causes cluster around leadership, data readiness, governance, and workflow redesign rather than model capability, and the organizations that avoid it consistently combine visible executive sponsorship with a defined, pre-agreed method for measuring adoption, quality, and financial outcome before they scale.

Introduction and Background

Enterprise deployment of artificial intelligence, and generative AI in particular, expanded rapidly between 2023 and 2026, but the share of that spending that reaches production with a measurable financial return has become one of the most contested figures in enterprise technology research. Independent studies from MIT, RAND, S&P Global Market Intelligence, and Gartner have each published a different estimate of how many AI initiatives fail, ranging from roughly 30% to more than 80%, and the gap between these numbers traces less to disagreement about facts than to differences in what each study measures: some count abandoned proof-of-concepts, others count production deployments that fail to reach a revenue or cost target, and others count organization-wide programs that never produced a measurable EBIT effect.

IntuitionLabs previously catalogued specific causes and case examples behind enterprise AI rollout failures in an April 2026 research report (intuitionlabs.ai). This report does not restate that catalogue. Instead, it assembles a source-linked evidence base built primarily from studies published or updated in 2025 and 2026, states the definitions and case-selection method behind each figure it cites, and lays the resulting numbers side by side in reproducible tables so a reader can judge which studies are measuring the same underlying phenomenon and which are not.

The report proceeds in five parts. It first defines the terms, pilot, deployment, adoption, quality, and financial outcome, that make cross-study comparison possible, and states the method used to select the studies and cases in this report. It then presents the documented failure and abandonment rate statistics from named research organizations, followed by a comparative taxonomy of the causes those organizations report. A dedicated section analyzes the pilot-to-production gap and the maturity models organizations use to track progress through it. A separate data section synthesizes the financial-outcome and ROI benchmarks published by major consulting and technology-research firms. A case-studies section then reviews documented, neutral examples of AI deployment in the life sciences industry specifically. Throughout, the report distinguishes a vendor's or company's own claim about its AI program from an independent survey finding, and an "announced" AI initiative from one that is "generally available" in production, because the two are conflated often enough in public discussion to distort the underlying failure-rate question.

Definitions and Case-Selection Method

Comparing failure-rate and ROI studies requires fixed definitions, because the studies surveyed for this report do not use identical ones. The National Institute of Standards and Technology (NIST) AI Risk Management Framework (AI RMF 1.0, January 2023) treats a **pilot** as one of several discrete "AI Deployment tasks" that precede full production, alongside checking compatibility with legacy systems and confirming regulatory compliance (nvlpubs.nist.gov). NIST's own risk-management lifecycle begins at a "Plan and Design" stage, since "risk management efforts start with the Plan and Design function in the application context," and runs through deployment (nvlpubs.nist.gov), meaning a **deployment**, in NIST's own usage, is the point at which a system moves from being tested against these tasks to being operated for its intended use. For judging whether that operation succeeds, NIST borrows the ISO 9000:2015 definition of **validation**: "confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled" (nvlpubs.nist.gov), and a related ISO/IEC definition of **reliability** as the "ability of an item to perform as required, without failure, for a given time interval, under given conditions" (nvlpubs.nist.gov), a component this report treats as part of deployment "quality" rather than adoption or financial outcome. At the organizational-management level, the International Organization for Standardization's ISO/IEC 42001:2023 standard formally "specifies requirements for establishing, implementing, maintaining, and continually improving an Artificial Intelligence Management System (AIMS)" (iso.org), the closest thing to a formal, auditable definition of what a well-governed AI deployment program looks like.

This report treats **adoption** and **financial outcome** as separate measurements from quality, following the practical distinction that IntuitionLabs, a life-sciences AI consultancy structured as an adjacent advisor rather than a software vendor, uses in its own evidence-measurement methodology: adoption metrics ("eligible users, activation, repeated use, workflow penetration, abandonment, template reuse, and participation by team or role") are tracked separately from outcome, quality-and-control, and sustainability metrics, because "the measure is not prompts sent," but

"whether an approved workflow becomes easier, faster, more reliable, or better supported without creating unacceptable risk or hidden work" (intuitionlabs.ai) (intuitionlabs.ai). That same methodology recommends agreeing success, stop, and review thresholds "before results are known where practical," specifically to avoid the temptation to redefine success around whatever a pilot happens to produce (intuitionlabs.ai).

The case-selection method for this report follows the same logic. Facts are included only when a named research organization, regulator, or company's own first-party material was directly retrieved and quoted during research for this report; secondary blogs, listicles, and content-farm restatements of a number are excluded even when they are the most visible source of that number in general search results. Every figure below states the originating organization, its publication date, and, where the source discloses it, its sample size and survey period, so that a reader can judge how much weight a given number should carry. Where two organizations' figures appear to conflict (for example, RAND's cited estimate that over 80% of AI projects fail against Gartner's 30% proof-of-concept abandonment prediction), both are reported with their original scope intact rather than collapsed into a single "AI failure rate."

Documented Failure and Abandonment Rates Across Major Studies

Table 1 below summarizes the headline failure and abandonment statistics from the primary studies located for this report, together with each study's stated method and sample.

Study (Originator)	Publication Date	Method / Sample	Headline Statistic
MIT NANDA, "The GenAI Divide: State of AI in Business 2025"	July 2025	Review of 300+ public AI initiatives; 52 structured interviews; 153 conference survey responses (research period January to June 2025)	95% of organizations piloting generative AI report zero measurable profit-and-loss return; only 5% extract significant value (nanda.media.mit.edu); of organizations evaluating enterprise-grade tools, only 20% reached pilot stage and 5% reached production (nanda.media.mit.edu)
RAND Corporation, "The Root Causes of Failure for Artificial Intelligence Projects"	August 13, 2024	65 interviews (50 industry practitioners, 15 academics), conducted August to December 2023	Cites outside estimates that more than 80% of AI projects fail, twice the failure rate of general information-technology projects (rand.org); RAND's own interviews found 84% of respondents cited leadership-related root causes as the primary reason AI projects fail (rand.org)
S&P Global Market Intelligence, Voice of the Enterprise: AI & Machine Learning, Use Cases 2025	May 30, 2025	Survey of 1,006 midlevel and senior IT and line-of-business professionals, North America and Europe	AI-initiative abandonment before production rose from 17% (2024) to 42% (2025) (spglobal.com); organizations report an average of 46% of AI projects scrapped between proof of concept and broad adoption (spglobal.com)
Gartner, press release	July 29, 2024	Analyst prediction informed by a survey of 822 business leaders fielded September to November 2023	At least 30% of generative AI projects predicted to be abandoned after proof of concept by the end of 2025, due to poor data quality, inadequate risk controls, escalating cost, or unclear business value (gartner.com)
Gartner, press release	June 25, 2025	Analyst prediction	Over 40% of agentic AI projects predicted to be canceled by the end of 2027, due to escalating cost, unclear business value, or inadequate risk controls (gartner.com)

Study (Originator)	Publication Date	Method / Sample	Headline Statistic
Forrester, Predictions 2025: Artificial Intelligence	2024	Analyst prediction	Three of four firms attempting to build their own advanced agentic AI architectures are predicted to fail (forrester.com)

These figures are not restatements of one another, and treating them as if they measure the same thing produces the misleadingly precise "X% of AI projects fail" framing common in secondary coverage. MIT NANDA measured realized profit-and-loss impact among organizations that had already piloted generative AI; RAND measured practitioners' own attribution of root cause across a broader, multi-year set of AI projects; S&P Global measured self-reported abandonment across the full initiative population, including projects that never reached a formal pilot; and Gartner's figures are forward-looking analyst predictions rather than retrospective survey findings. Gartner itself attributed its 2024 prediction partly to organizational economics rather than technical failure: its analysts noted that "executives are impatient to see returns on GenAI investments, yet organizations are struggling to prove and realize value" (gartner.com), a funding-model explanation rather than a claim about model quality.

Causes of Enterprise AI Implementation Failure: A Comparative Taxonomy

Across the six research organizations reviewed for this report, the documented causes of enterprise AI implementation failure cluster into six recurring categories rather than a single technical explanation. *Table 2* below compares how each organization frames these causes.

Cause Category	Selected Findings by Source
Data quality and readiness	Poor data quality is one of Gartner's top cited reasons for GenAI project abandonment, discussed above. IBM states that "poor-quality data weakens the performance and reliability of AI models" (ibm.com). Deloitte finds that "legacy data and infrastructure architectures cannot power real-time, autonomous AI" (deloitte.com).
Unclear ROI and business value	McKinsey's November 2025 survey found only 37% of respondents report AI has positively contributed to EBIT, "essentially unchanged from 2025" (mckinsey.com). BCG cites "missing clear AI metrics and ROI measurement" among the barriers limiting value capture (bcg.com).
Talent and skills gaps	Deloitte's 2026 enterprise survey finds "insufficient worker skills are the biggest barrier to integrating AI into existing workflows" (deloitte.com). BCG's 2025 study lists "shortage of AI talent" among the top-cited scaling barriers (bcg.com).
Governance and risk controls	Inadequate risk controls are also among Gartner's cited drivers of GenAI abandonment, discussed above. IBM notes companies "often lack policies for monitoring AI behavior, reviewing automated decisions or assigning accountability" (ibm.com). Deloitte finds only one in five companies has "a mature model for governance of autonomous AI agents" (deloitte.com).
Leadership and organizational commitment	RAND's practitioner interviews found leadership-related causes, not technical failure, drove 84% of cited failure reasons (rand.org). McKinsey finds high performers are "twice as likely as others" to report senior-leader commitment and defined impact-measurement processes (mckinsey.com). IBM frames the core gap as organizational: "AI capability is advancing faster than organizational capability" (ibm.com). BCG finds "senior executives in future-built organizations act as highly visible sponsors of AI" (bcg.com).

Cause Category	Selected Findings by Source
Integration and workflow redesign	McKinsey finds high performers "fundamentally redesign workflows that are enabled by AI rather than insert AI into existing ones" (mckinsey.com). IBM notes integration problems "can limit system performance, create bottlenecks or introduce security risks" (ibm.com). BCG states that "integrating agents into a workflow is not a matter of bolting them onto legacy processes" but requires end-to-end redesign (bcg.com).

The pattern across all six sources is that failure is rarely attributed to the underlying model failing to work as designed; it is attributed to the surrounding organizational system, funding model, governance structure, or workflow design failing to change enough to use it.

Analysis of Key Segments: The Pilot-to-Production Gap and Adoption Maturity

The gap between an AI pilot and a scaled production deployment is itself a measurable, well-documented segment of enterprise AI outcomes, distinct from either the failure-rate or ROI questions addressed elsewhere in this report. O'Reilly's November 2023 survey found that 67% of organizations reported already using generative AI (oreilly.com), yet only a minority of those organizations had progressed past experimentation, and just 18% reported having AI applications already in production (oreilly.com). The same survey identified the leading barrier to advancing past the pilot stage as identifying appropriate use cases, cited by 53% of respondents (oreilly.com), ahead of legal, risk, and compliance concerns. An IDC survey of 900 organizations across Asia-Pacific, commissioned by Lenovo, found the drop-off even steeper: "less than 10% of total POCs were actually deemed successful, having met predefined business goals" (lenovo.com), a figure that is not directly comparable to Gartner's 30% proof-of-concept abandonment prediction because it measures success against predefined goals rather than simple continuation past the pilot stage. Forrester's Predictions 2025 report adds a related, forward-looking data point specific to agentic AI: it forecasts that "three out of four firms that build aspirational agentic architectures on their own will fail" (forrester.com), suggesting the pilot-to-production gap may widen further as enterprises move from single-model generative AI to multi-step autonomous agent systems.

Maturity models offer one framework for locating an organization within this gap. Google Cloud's AI Adoption Framework defines "three natural phases to AI maturity: tactical, strategic, and transformational" (google.com), where the tactical phase is characterized by narrow, short-term use cases, and the strategic phase is defined by "several ML systems deployed and maintained in production" delivering "sustainable business value" (google.com). Deloitte's parallel maturity research classifies organizations into "Automators," using single-agent workflows for basic automation (deloitte.com), and more advanced "Transformers," and finds the more mature group applies a broader financial, customer, process, and workforce key-performance-indicator (KPI) set more consistently: "73% of Transformers surveyed reported using all 46 KPIs frequently or very frequently, compared with 69% of Automators" (deloitte.com). MIT Sloan Management Review's own analysis of ROI-measurement practice adds an important qualifier: measuring return differs by AI type, since generative AI improvements in speed, quality, or output volume require "deliberate translation into financial impact" that analytical AI's more directly attributable gains do not (sloanreview.mit.edu).

A further, underdocumented dimension of the pilot-to-production gap is workforce-level adoption after a system is technically in production. BCG's 2025 "AI at Work" study found that regular generative AI use "has stalled at 51%" among frontline employees even as leadership usage is higher (bcg.com), a gap BCG attributes substantially to

leadership behavior: the share of employees who feel positive about generative AI "rises from 15% to 55% with strong leadership support" (bcg.com). A separate BCG study found "more than 85% of employees remain at stages two and three" of a four-stage AI adoption curve rather than the more advanced stages associated with material value (bcg.com), and correspondingly found that "60% of companies globally were not generating any material value from AI despite substantial investment" (bcg.com). Microsoft's own 2024 Work Trend Index adds a governance-relevant data point: while 79% of leaders agreed their company needed to adopt AI to stay competitive (microsoft.com), 78% of AI users reported "bringing their own AI tools to work" outside any official deployment (microsoft.com), meaning a meaningful share of "AI adoption" in practice occurs outside any deployment an enterprise could measure, govern, or count as a supported production system. IntuitionLabs' own advisory approach reflects this segment explicitly, structuring engagements "one department at a time" with "governed information, specialist implementation, role-based adoption, and measured results" rather than an enterprise-wide rollout (intuitionlabs.ai), a scoping choice consistent with the pilot-to-production evidence above, in which enterprise-wide, unscoped rollouts appear to correlate with the higher abandonment figures reported by S&P Global and IDC.

Data Analysis and Evidence

Financial-outcome data from major research and consulting firms present a consistent pattern across independent methodologies: a small minority of enterprises report significant realized value from AI, most report modest or no measurable value, and the gap between the two groups is widening rather than narrowing between 2023 and 2026. *Table 3* below compares the largest published studies.

Research Firm / Study	Date	Sample	Key Financial-Outcome Finding
McKinsey, The State of AI in 2025	November 2025	n=1,993	39% attribute any level of EBIT impact to AI enterprise-wide; roughly 6% qualify as "AI high performers" reporting 5%+ EBIT impact (mckinsey.com)
McKinsey, The State of AI (2025 edition)	March 2025	n=1,491, 101 countries	17% report 5% or more of enterprise EBIT attributable to gen AI; over 80% report no tangible enterprise-level EBIT impact (mckinsey.com)
IBM Institute for Business Value	June 2023	n=2,500 executives, 34 roles, 16 countries	Average enterprise-wide AI ROI of just 5.9%, below the typical 10% cost of capital; fewer than one in four organizations exceed 10% ROI (ibm.com)
IBM / Oxford Economics, global CEO study	May 2025	n=2,000 CEOs, 33 countries	Only 25% of AI initiatives have delivered expected ROI; only 16% have scaled enterprise-wide (ibm.com); 85% expect scaled AI efficiency investments to return positive ROI by 2027 (ibm.com)
Deloitte, AI ROI: The Paradox of Rising Investment	October 2025	n=1,854 executives	Only about 1 in 5 organizations qualify as "AI ROI Leaders" (deloitte.com); 15% of gen AI users already report significant, measurable ROI, versus just 10% for agentic AI specifically (deloitte.com)
BCG, The Widening AI Value Gap	September 2025	n=1,250 executives, 9 industries	5% of companies are "future-built" for AI, achieving 1.7x revenue growth and 3.6x three-year total shareholder return versus laggards (bcg.com); 60% are "laggards" reporting minimal gains

Research Firm / Study	Date	Sample	Key Financial-Outcome Finding
Accenture, Making Reinvention Real with Gen AI	March 2025	3,000+ C-level executives surveyed; 2,000+ client projects reviewed	Only 36% of executives have scaled gen AI solutions; just 13% report significant enterprise-level value (accenture.com); executive buy-in correlates with 2.5x higher ROI (accenture.com)
Wharton / GBK Collective, AI Adoption Report	October 2025	n=801 business leaders (2025 wave)	72% now track structured, business-linked ROI metrics, up from prior years (wharton.upenn.edu); nearly three-quarters already report positive ROI, and four in five expect positive returns within two to three years (wharton.upenn.edu)

The pattern across these eight independent studies is directionally consistent even though the exact percentages differ by definition and sample: a small, single-digit-to-low-double-digit share of enterprises report scaled, significant financial value from AI (McKinsey's roughly 6%, BCG's 5%, Accenture's 13%, Deloitte's 10% for agentic AI specifically), while a much larger share, generally 60% to 80% depending on the study, report modest, unmeasured, or negligible impact at the enterprise level. Notably, the Wharton/GBK Collective figures are directionally more optimistic than McKinsey's or IBM's, which this report attributes at least partly to a definitional difference: Wharton and GBK's survey measures leaders' self-reported perception of ROI, including expectations of future returns, while McKinsey and IBM measure realized, enterprise-level EBIT or ROI against a specific financial threshold. This distinction, between a leader's sentiment about value and a measured financial outcome, is itself one of the primary sources of disagreement in public discussion of enterprise AI's return on investment, and readers comparing figures across studies should check which of the two each source is actually reporting.

Case Studies and Real-World Examples

The following cases illustrate documented, first-party or peer-reviewed AI deployments in life sciences specifically, reported here in neutral, descriptive terms as documented facts about scope, timing, and stated outcome rather than as endorsements of a vendor's claims.

Sanofi: Company-Wide AI Deployment Announcement (2023)

In June 2023, Sanofi announced a company-wide AI initiative, with its chief executive describing an ambition to become "the first pharma company powered by artificial intelligence at scale" (sanofi.com). The company's own release stated that AI tools had cut some research-process timelines "from a matter of weeks to just hours" in target-identification work (sanofi.com), and reported that its internal "plai" application, once adopted in its biopharmaceutical supply chain, could predict 80% of low-inventory positions (sanofi.com). These figures are vendor-reported, first-party claims rather than independently audited outcomes, and this report presents them on that basis.

Novartis and Microsoft: Multi-Year Drug-Discovery Collaboration (2019 onward)

Novartis and Microsoft announced a multi-year collaboration in October 2019 applying AI, including generative-chemistry methods, across drug discovery and development, in a release titled "Novartis and Microsoft announce collaboration to transform medicine with artificial intelligence" (novartis.com). The collaboration is a documented

example of an announced, multi-year enterprise AI program in life sciences research and development rather than a short-lived pilot.

Insilico Medicine: AI-Designed Molecule Reaches Clinical Data (2025)

Insilico Medicine describes rentosertib as a "first-in-class small molecule targeting TNIK developed utilizing generative AI" (insilico.com). According to the company's own June 2025 release describing results it says were published in Nature Medicine, its Phase IIa trial's high-dose arm showed a mean forced vital capacity (FVC) increase of +98.4 mL relative to the comparison group (insilico.com). This case illustrates a generative-AI-originated drug candidate reaching a mid-stage clinical readout, distinct from the deployment and adoption metrics discussed elsewhere in this report, since it measures a clinical outcome rather than an enterprise-operations outcome.

US Health Systems: Documented AI Piloting and Barriers (2024 to 2025)

A peer-reviewed survey of members of the Scottsdale Institute, a network of non-profit US health systems, was fielded in fall 2024 and included 67 health systems with a 64% response rate (pmc.ncbi.nlm.nih.gov). The study found that immature AI tooling was the most commonly cited barrier to deployment (pmc.ncbi.nlm.nih.gov), even though every responding health system had "at least begun developing or piloting" ambient clinical-documentation AI specifically (pmc.ncbi.nlm.nih.gov), illustrating the announced-versus-scaled distinction discussed earlier in this report within a single clinical use case. Separately, Deloitte's survey of 280 life-sciences C-suite executives found that only a minority reported having "successfully scaled AI, and just 9% reported achieving" its full expected benefit (deloitte.com).

DIA and Tufts CSDD: Multi-Company Clinical Development Study (2024)

The Drug Information Association (DIA) and the Tufts Center for the Study of Drug Development launched a study involving 16 biopharmaceutical companies and contract research organizations (CROs) examining AI, machine learning, and natural language processing use across specific clinical-development functions, including "site identification, patient recruitment, pharmacovigilance, quality assurance, and clinical monitoring" (diaglobal.org). The FDA separately maintains a public list of AI-enabled medical devices it has authorized for the US market, which the agency itself describes as "a resource intended to identify AI-enabled medical devices that are authorized" rather than an exhaustive registry of every AI tool in clinical use (fda.gov), a distinction relevant to any reader comparing "how many AI tools are approved" against "how many are in generally available production use."

Implications and Future Directions

Several implications follow from placing these studies side by side rather than treating any single failure-rate or ROI figure as definitive. First, the wide spread in headline failure rates, from Gartner's 30% proof-of-concept abandonment prediction to RAND's cited 80%-plus estimate, is substantially explained by differences in what is being measured, and future reporting on this topic should specify whether a cited figure describes proof-of-concept abandonment, production-deployment failure, or enterprise-level value realization, since these are three different populations with three different denominators. Second, the causal evidence assembled in this report's taxonomy points consistently toward organizational and governance factors, not model capability, as the dominant explanation

for failure, a conclusion reinforced by RAND's finding that 84% of practitioner-cited causes were leadership-related and by McKinsey's, IBM's, Deloitte's, and BCG's independent findings that data readiness, governance maturity, and executive sponsorship separate high performers from the rest.

Third, the gap between announced AI capability and generally available, production-scale deployment remains wide across every study reviewed, from O'Reilly's 18% production rate against 67% generative AI usage, to IDC's finding that fewer than 10% of proof-of-concepts meet predefined business goals, to Deloitte's finding that only one in five companies has mature governance for autonomous agents even as agentic AI adoption is expected to grow. This gap is likely to widen further before it narrows: Gartner's prediction that over 40% of agentic AI projects will be canceled by 2027 and Forrester's prediction that three of four self-built agentic architectures will fail both suggest that the added complexity of multi-step, autonomous AI systems compounds the organizational failure modes already documented for single-model generative AI, rather than resolving them.

Fourth, life sciences organizations face this gap alongside sector-specific constraints, including regulatory validation requirements and the distinction between an FDA-authorized AI-enabled device and one in routine production use, that do not appear in general enterprise AI research. The evidence in this report, spanning a company-wide vendor announcement (Sanofi), a multi-year research collaboration (Novartis and Microsoft), a generative-AI-originated drug candidate reaching mid-stage clinical data (Insilico Medicine), and independent surveys of health systems and life-sciences executives, suggests that the life sciences sector's AI deployment pattern broadly mirrors the general-enterprise pattern documented earlier in this report: widespread piloting, a minority reaching full production scale, and a further minority reporting the full expected financial or operational return. For organizations planning AI deployments in this sector, the definitions, method, and comparative evidence set out above provide a basis for setting realistic, pre-agreed success thresholds rather than adopting a headline failure-rate or ROI figure that may describe a different population than the one they are planning for.

Conclusion

The evidence assembled in this report does not support a single, precise "enterprise AI failure rate," because the studies that produce these figures measure different populations, different stages, and different definitions of success. What it does support is a consistent, cross-validated pattern: a small minority of enterprises, generally in the single digits to low teens depending on the study, report scaled AI deployments with measurable, significant financial return, a much larger share report modest or negligible enterprise-level impact despite substantial investment, and the organizations that separate themselves from this pattern do so through documented leadership commitment, data and governance readiness, and workflow redesign rather than through access to more capable AI models. The pilot-to-production gap is itself a measurable and largely organizational phenomenon, not primarily a technical one, and it appears to widen rather than narrow as enterprises move from single-model generative AI toward multi-step agentic AI systems. Life sciences organizations face this same general pattern with sector-specific regulatory and validation constraints layered on top. Readers using any of the statistics in this report, or encountering similar statistics elsewhere, should ask what population, stage, and definition of success the figure describes before treating it as comparable to any other figure in the same conversation.

Frequently Asked Questions (FAQs)

What percentage of enterprise AI projects fail? There is no single agreed figure, because studies measure different things. As Table 1 above details, published estimates range from Gartner's prediction that 30% of generative AI projects would be abandoned after proof of concept by the end of 2025 (gartner.com) to RAND's cited estimate that over 80% of AI projects fail (rand.org), depending on the definition and stage each study measures.

Why do enterprise AI projects fail? The dominant documented causes are organizational rather than technical: data readiness, unclear ROI measurement, talent gaps, governance and risk-control maturity, leadership commitment, and workflow redesign, as detailed in this report's comparative taxonomy above.

What is the difference between an AI pilot and a production AI deployment? NIST's AI Risk Management Framework treats piloting as one of several discrete tasks that precede full deployment, alongside legacy-system compatibility checks and regulatory-compliance confirmation (nvlpubs.nist.gov); a production deployment is the point at which the system is operated for its intended, validated use.

What ROI can enterprises expect from AI? Published findings vary by study and definition. IBM's Institute for Business Value found an average enterprise AI ROI of 5.9% (ibm.com), while McKinsey found only 39% of respondents attribute any EBIT impact to AI at all (mckinsey.com).

How is AI deployment success measured? This report distinguishes adoption metrics (usage, activation, workflow penetration), quality metrics (accuracy, reliability, exceptions), and financial-outcome metrics (ROI, EBIT impact, cost savings), following the definitional framework set out in the Definitions and Case-Selection Method section above.

Are there documented AI deployment case studies in life sciences? Yes. This report documents Sanofi's 2023 company-wide AI announcement (sanofi.com), the Novartis-Microsoft drug-discovery collaboration (novartis.com), Insilico Medicine's AI-designed clinical candidate (insilico.com), and independent surveys of US health systems and life-sciences executives, all detailed in the Case Studies section above.
