

INFOGRAPHIC · AGENTIC AI · LINKEDIN

Cancellation of Agentic AI Projects by 2027: An Analysis of Gartner's Prediction

What the forecast claims, what the literature confirms and how to change the trajectory

Aridio Silva - Independent Researcher · @aridiosilva · ORCID 0009-0008-2411-6995

© 2026 Aridio Silva @aridiosilva · ORCID 0009-0008-2411-6995 · October 2026

Five numbers to understand the debate

40%+

of agentic AI projects will be canceled by the end of 2027 (Gartner forecast, Jun. 2025)

17%

of organizations have already deployed agents; over 60% expect to within 2 years (Gartner, Apr. 2026)

37%

attribute some EBIT to AI in 2026, versus 39% in 2025 (McKinsey)

14% | 32%

cut headcount because of AI (2026) | expected to (2025). Inflated expectations (McKinsey)

40.6%

was ChatDev's best accuracy after design tweaks (it started at 25.0%); MAST study

★ Importance

Sums up the debate: a cancellation forecast coexists with growing adoption and uneven returns.

+ Positive implications

The field is not collapsing: use of agents at scale rose among large companies.

- Negative implications

The numbers come from different sources and methods and cannot be added together.

Research question and how we checked the data

- **Question:** why is a significant share of agentic AI projects expected to be canceled, and how can this risk be reduced?
- **When:** Oct 14, 2026, through web search and direct reading of the original pages.
- **First link:** share.google would not open (rate limit); the primary source was located by search.
- **Limit:** Gartner's paid report was not consulted; the analysis uses the public press release.
- **Criterion:** each number checked against a source, with an explicit verdict.

Importance

Explains how the numbers were checked and what was left out.

Positive implications

Transparency lets the reader redo each check.

Negative implications

Paid and secondary sources limit part of the conclusions.

The four verdicts

 Holds

 Holds in part

 Does not hold

 Not verified

Verification status of each source

Source	Verification status
Gartner (2025a), the 40% press release	Read in full; paid report not accessed
Gartner (2024), 30% of generative AI	Checked in the opening text
Gartner (2025b), AI Hype Cycle	Public abstract only; positions via secondary coverage
Gartner (2026), Agentic AI Hype Cycle	Read in full
RAND (2024), RR-A2680-1	Read in full (original)
McKinsey (2025) and (2026), State of AI	Read in full (originals)
Cemri et al. (2025), MAST	Read in full (arXiv v3)
MIT NANDA (2025)	Original not accessed; copy hosted by a third party
τ-bench and TheAgentCompany	Official abstracts only (arXiv and ICLR)
Lighthill (1973) and the "AI winters"	Secondary source (DataCamp); recheck

Green: original read · Amber: abstract or opening text · Coral: secondary source or not accessed

Gartner's press release of June 25, 2025

40%+

of agentic AI projects will be canceled by the end of 2027

- **Origin:** press release from Sydney, with statements by senior director analyst Anushree Verma.
- **Reasons:** escalating costs, unclear business value or inadequate risk controls.
- **Basis:** paid report "Avoid Agentic AI Failure"; the calculation method is not disclosed in the press release.
- **Nature:** analyst forecast, not a scientific paper.

★ Importance

It is the origin of the headline; the whole debate starts from this press release.

+ Positive implications

Warns early about cost, value and risk before scaling a project.

− Negative implications

A number without a public method can be repeated as if it had been measured.

The press release's four claims

1 · Cancellation

Over 40% of agentic projects will be canceled by the end of 2027.

2 · Hype

Most current projects are experiments or proofs of concept, driven by hype and often misapplied.

3 · Agent washing

Vendors rebrand assistants, RPA and chatbots as agents; only about 130 of the thousands of vendors are real.

4 · Growth

By 2028, at least 15% of day-to-day decisions will be autonomous (0% in 2024) and 33% of enterprise software will include agentic AI.

Importance

Shows that the forecast is about projects, not against the technology.

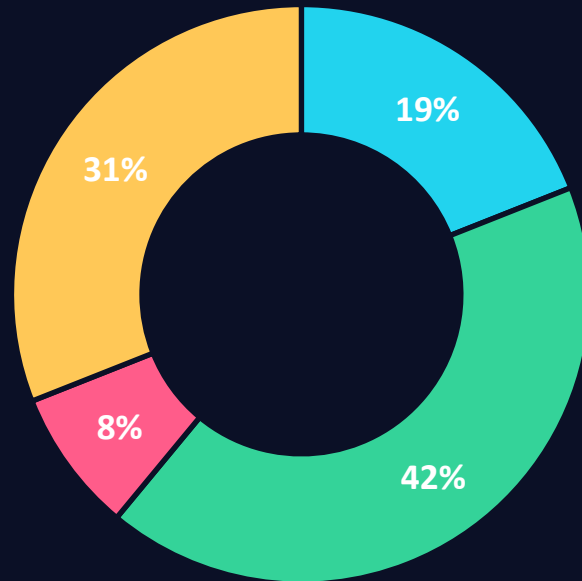
Positive implications

Gartner itself projects growing adoption through 2028.

Negative implications

Hype and agent washing inflate the number of poorly grounded projects.

The January 2025 poll measures investment intent



- Significant investment
- Conservative investment
- No investment
- Waiting or unsure

- **Basis:** 3,412 Gartner webinar attendees, in January 2025.
- **Reading:** 61% already invested (19% significantly and 42% conservatively).
- **Caution:** the poll is not the stated basis of the 40% figure.

★ Importance

Avoids linking the 40% to a poll that measures something else.

+ Positive implications

Shows real interest: 61% already invested in agentic AI.

- Negative implications

Almost a third (31%) is still waiting or unsure.

Core concepts

Agentic AI

Systems that autonomously plan and take actions to meet user-defined goals (Gartner, 2024).

AI agent

LLM-based entity with instructions, a conversation history and the ability to act on its environment, for example by using tools (Cemri et al., 2025).

Autonomous AI

Headline term. In practice, agents that pursue a goal with little supervision, like AutoGPT in 2023.

Agent washing

Rebranding of assistants, RPA and chatbots as "agents" without real autonomy (Gartner, 2025a).

Pass^k

Probability that the agent succeeds on all k runs of the same task. Measures reliability, not just average accuracy (Yao et al., 2024).

Pilot-to-production gap

Distance between performance in the pilot (narrow scope, clean data, heavy supervision) and in production (McKinsey, 2025).

Metrics, standards and protocols

 Hype cycle

Gartner model: expectations rise to a peak, fall into the trough of disillusionment and then stabilize.

 EBIT attributable to AI

Share of operating profit that a company attributes to its use of AI; a value-at-scale indicator (McKinsey).

 Multi-agent system

A collection of agents designed to interact through orchestration (Cemri et al., 2025).

 MAST

Taxonomy of 14 failure modes in multi-agent systems, in 3 categories (Cemri et al., 2025).

 MCP

Model Context Protocol: Anthropic's open standard (Nov. 2024) for connecting models to tools and data.

 A2A

Agent2Agent: Google's open protocol (Apr. 2025) for agents from different vendors to communicate.

Acronyms and supporting terms

FinOps for agentic AI

Managing the cost of agents; a category that appears in the 2026 Hype Cycle (Gartner).

High performers

Organizations that McKinsey classifies as high performers with AI.

PoC

Proof of concept: a limited test to check whether an idea is viable.

ROI

Return on investment: the gain obtained relative to the cost.

RPA

Robotic process automation: software robots that follow fixed rules.

LLM

Large language model, the basis of most current agents.

GRC

Governance, risk and compliance: the discipline that defines and verifies controls.

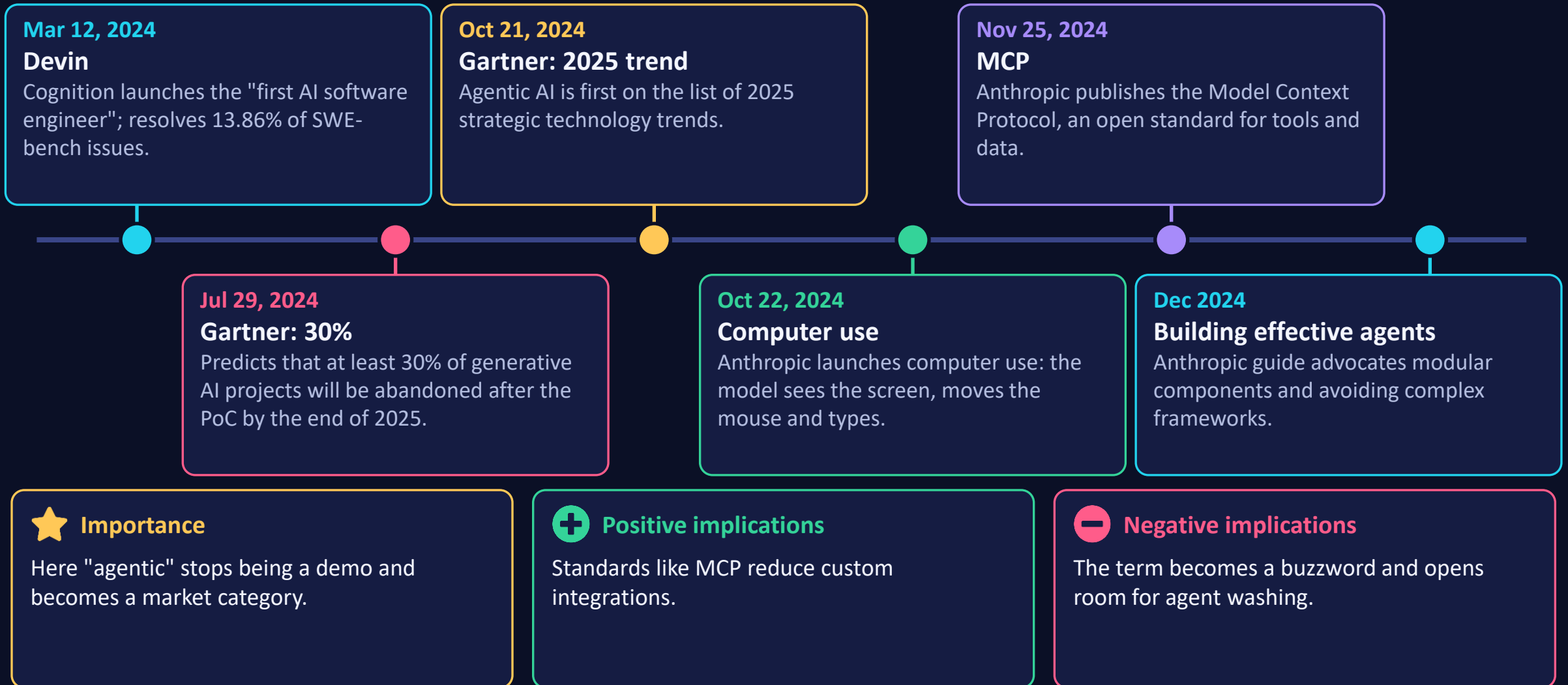
Benchmark

Standardized set of tasks for comparing systems (e.g., τ -bench).

2022–2023: from conversation to agent



2024: "agentic" becomes a market category



2025: protocols, hype and first warnings



2026: peak of expectations and adjustment

Apr 15, 2026

Gartner: Agentic AI Hype Cycle

Agentic AI at the peak; only 17% have deployed agents; governance, security and FinOps emerge as categories.

Aug 25, 2026

McKinsey 2026

40% of large companies scale agents (27% in 2025); attributed EBIT stable at 37%.

Apr 2026

A2A reaches version 1.0

Protocol gathers over 150 supporting organizations, according to O'Reilly Radar.

★ Importance

Shows the present: adoption grows, but maturity has not kept pace.

+ Positive implications

Governance and FinOps for agents become categories of their own.

− Negative implications

Being at the peak of expectations signals a risk of another correction.

Variants of the concept and when they emerged

Term or variant	Emerges with	When	Source
Autonomous agent, "autonomous AI"	AutoGPT	Mar 30, 2023	Wikipedia; press (secondary)
Multi-agent system	ChatDev, MetaGPT, AutoGen	Jul–Aug 2023	Reference list of Cemri et al.
Software agent	Devin (Cognition)	Mar 12, 2024	Cognition via press (secondary)
Agentic AI (category)	2025 strategic trends	Oct 21, 2024	Gartner (press release)
Computer use	Anthropic	Oct 22, 2024	Technical blog (secondary)
Agent protocols	MCP and A2A	Nov 2024 and Apr 2025	Press and IBM (secondary)
Agent washing	The 40% press release	Jun 25, 2025	Gartner (2025a)
Governance, security and FinOps for agents	Agentic AI Hype Cycle	Apr 2026	Gartner (2026)

Author's reading: "autonomous AI" (2023) and "agentic AI" (2024) name the same idea at different moments: software that plans and acts to meet a goal.

The "AI winters" already showed this pattern

1973–1974 · Lighthill report

Criticized AI research and contributed to DARPA's funding cut in 1974.

1980s · Expert systems

MYCIN and XCON attract investment; Japan's Fifth Generation project targets computers for them.

1987 · Market collapse

The market for expert systems and Lisp machines collapses: maintenance cost more than the value delivered.

Author's reading

Similarity: the value of isolated cases did not survive the cost of maintenance at scale.

Difference: in 2026 use is broad and growing; 44% of organizations already scale AI across the enterprise (McKinsey, 2026).

★ Importance

Shows that cycles of promise and frustration have already occurred in AI history.

+ Positive implications

The field reorganized and came back: a correction does not mean the end.

− Negative implications

In 1987 maintenance cost exceeded value: a similar risk today.



Rising cost and unclear value

- **Gartner:** escalating costs and uncertain value are central reasons for cancellation.
- **Gartner:** many "agentic" use cases could be solved with simpler tools.
- **Gartner:** integrating agents into legacy systems is technically complex and requires expensive modifications.
- **RAND:** leaders sometimes demand AI for problems that a few simple rules would solve.



★ Importance

It is the central reason cited in Gartner's press release.

+ Positive implications

Choosing the right case and the simplest tool reduces cost.

- Negative implications

A project without clear ROI is the first to be cut.

Technical reliability of agents

$\approx 30\%$

of tasks completed by the best agent in
TheAgentCompany (abstract of the
revised version)

$< 50\%$

of tasks solved by state-of-the-art agents
on τ -bench

$< 25\%$

success at pass⁸ in retail (τ -bench):
inconsistency across repeated runs

Caution: these are 2024 benchmarks. They serve as evidence of the type of failure (inconsistency, long tasks), not as a current scoreboard.

★ Importance

Agents must be right every time, not just on average.

+ Positive implications

Measuring pass^k exposes inconsistency before production.

- Negative implications

An agent that is right once and wrong on the repeat breaks real processes.

System design failures (MAST)

1.642

annotated runs of 7 systems

14

failure modes in 3 categories

41–86.7%

failure rates across the 7 systems

*Authors' thesis (conjecture):
improving the base model alone is
not enough.*

- **System design:** poorly defined roles, instructions and termination criteria.
- **Misalignment between agents:** information that does not flow, reasoning that does not become action.
- **Task verification:** missing, incomplete or incorrect checking.

★ Importance

Maps where multi-agent systems fail.

+ Positive implications

Design failures can be fixed with architecture and verification.

– Negative implications

Even with tweaks, accuracy stayed below half in the tested case.

Ill-defined problem and leadership

84%

of interviewees cited at least one leadership-related root cause (RAND, 2024)

- **Wrong metric:** optimizing for something that is not the business goal.
- **Timing:** underestimating the time the project requires.
- **Priority:** frequent changes of focus interrupt the project.
- **Application:** using AI where simple rules would do.

Importance

It is the most cited cause in the RAND interviews.

Positive implications

Aligning business and technical teams before the pilot is cheap and effective.

Negative implications

A sample mostly of engineers may bias results toward leadership failures.



Insufficient data and infrastructure

30 of 50

industry interviewees reported persistent data quality problems (RAND, 2024)

- **Data:** quality and fit to the problem appear among the five root causes.
- **Infrastructure:** lack of data and deployment infrastructure blocks the move to production.
- **RAND limit:** the study excludes projects using only pretrained LLMs and its interviews are from 2023.

★ Importance

Without adequate data and infrastructure, the model does not reach production.

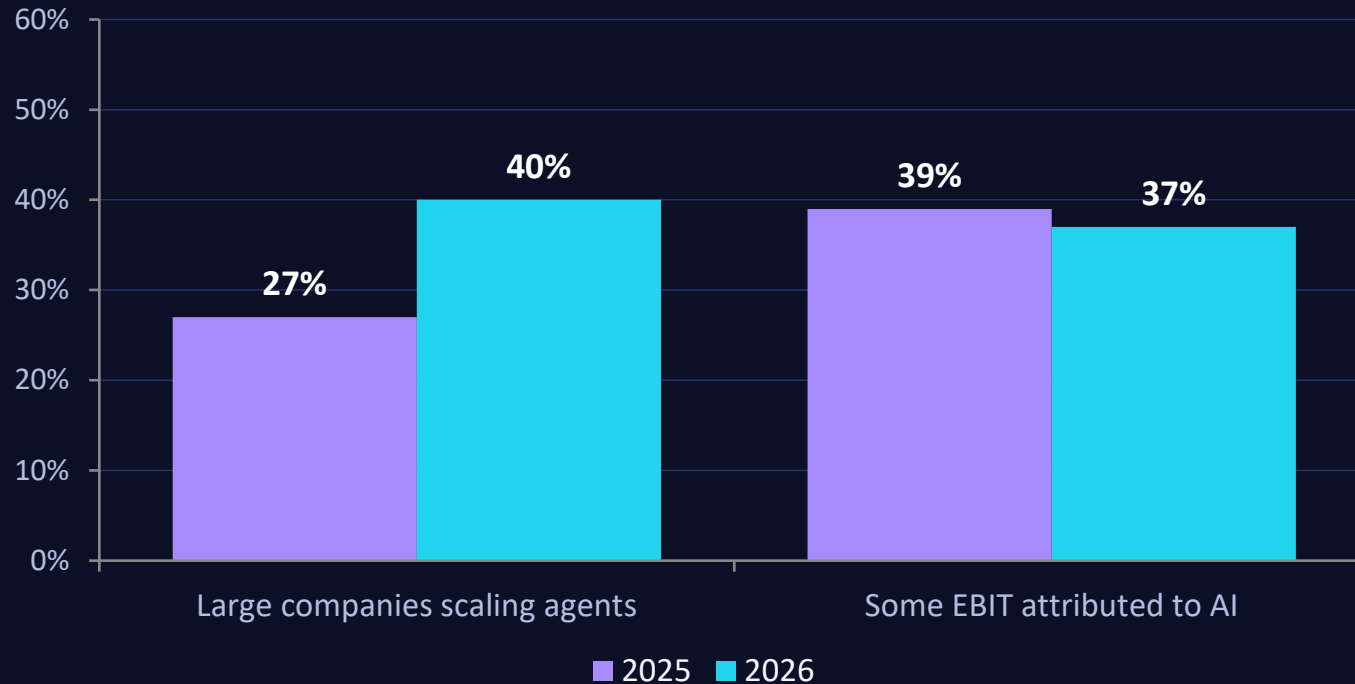
+ Positive implications

Investing in data and deployment before scaling reduces rework.

- Negative implications

Data cost and timelines are usually underestimated.

Scale and financial return move at different speeds



- **Scale:** large companies scaling agents rose from 27% to 40%.
- **Return:** EBIT attributed to AI stayed flat (39% and 37%).
- **Cost:** about 20% said operating costs, including tokens, limited use.

★ Importance

Shows the distance between use and financial return.

+ Positive implications

Scale at large companies rose from 27% to 40%.

- Negative implications

Attributed EBIT stayed flat and about 20% cite costs as a brake.

Hype and agent washing

~130

real vendors among the thousands that call themselves agentic (Gartner, 2025a)

2025 and 2026

Agentic AI at the peak of inflated expectations in Gartner's cycle

★ Importance

Inflated expectations push the purchase of "agents" that are not agents.

+ Positive implications

Demanding proof of real autonomy filters out vendors.

- Negative implications

Buying a rebranded chatbot produces a project with no value.

- **Rebranding:** assistants, RPA and chatbots sold as agents.
- **Gap:** Gartner describes a growing gap between ambition and execution (2026).
- **Scope:** most deployments remain narrowly scoped.

Five caveats when interpreting the evidence

- **2024 benchmarks:** models already superseded; they show the type of failure, not the current scoreboard.
- **RAND:** excludes projects using only pretrained LLMs; interviews done in 2023.
- **RAND sample:** mostly engineers without management roles; bias toward leadership failures.
- **McKinsey:** self-reported surveys, weighted by GDP; they measure scale, not failure.
- **MIT NANDA:** preliminary report with 153 respondents, 52 interviews and 300 cases; original not checked.



★ Importance

Avoids treating each number as definitive proof.

+ Positive implications

Critical readers extract more value from each source.

- Negative implications

Part of the evidence is indirect or ages quickly.



Canceling is not the same as failing technically

- **Priorities:** leaders sometimes change focus every few weeks or months.
- **Discarding:** projects under way are abandoned before showing results, or completed and then ignored.
- **Probabilistic:** models do not always repeat the same result; those who expect repeatability lose confidence.
- **Link:** an organizational mechanism consistent with "unclear business value" in Gartner's press release.



★ Importance

Explains why "canceled" does not prove that the technology failed.

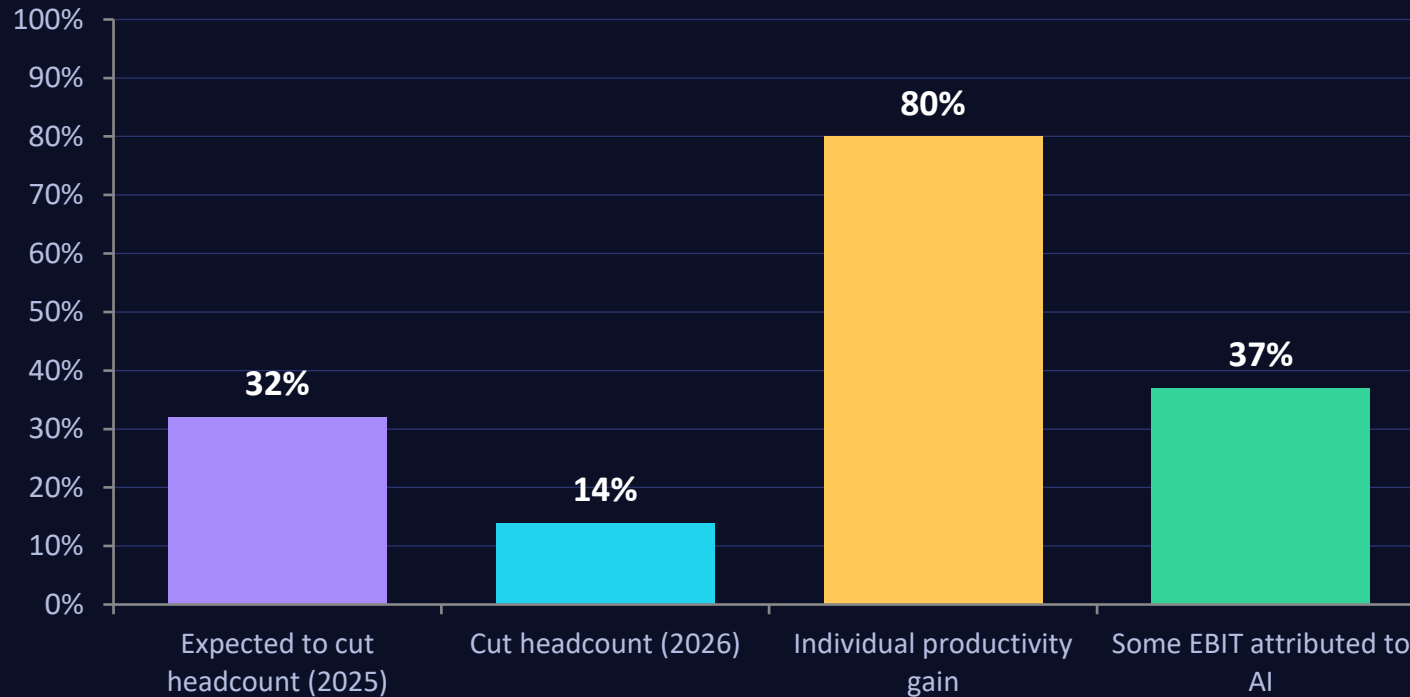
+ Positive implications

Stable portfolio governance protects good projects.

− Negative implications

Priority changes destroy value before it appears.

Inflated expectations were measured



- **Headcount:** 14% report AI-related reductions, less than half of the 32% who expected them in 2025.
- **Individual versus enterprise:** 80% say AI improved their own productivity, but only 37% attribute EBIT to it.

★ Importance

Measures the hype: the expected headcount cut was more than double what was observed.

+ Positive implications

Individual productivity gain is widespread (80%).

− Negative implications

Individual gain rarely reaches the bottom line (37%).

Risk and cost grow with scale

51%

of organizations using AI reported at least one negative consequence (2025)

~1/3

cited inaccuracy as a negative consequence (2025)

≈3×

more constrained by cost in using coding agents: the high performers (2026)

Explainability was the second most reported risk but is not among the most mitigated (McKinsey, 2025).

★ Importance

Shows that mitigating risk and controlling cost are permanent tasks.

+ Positive implications

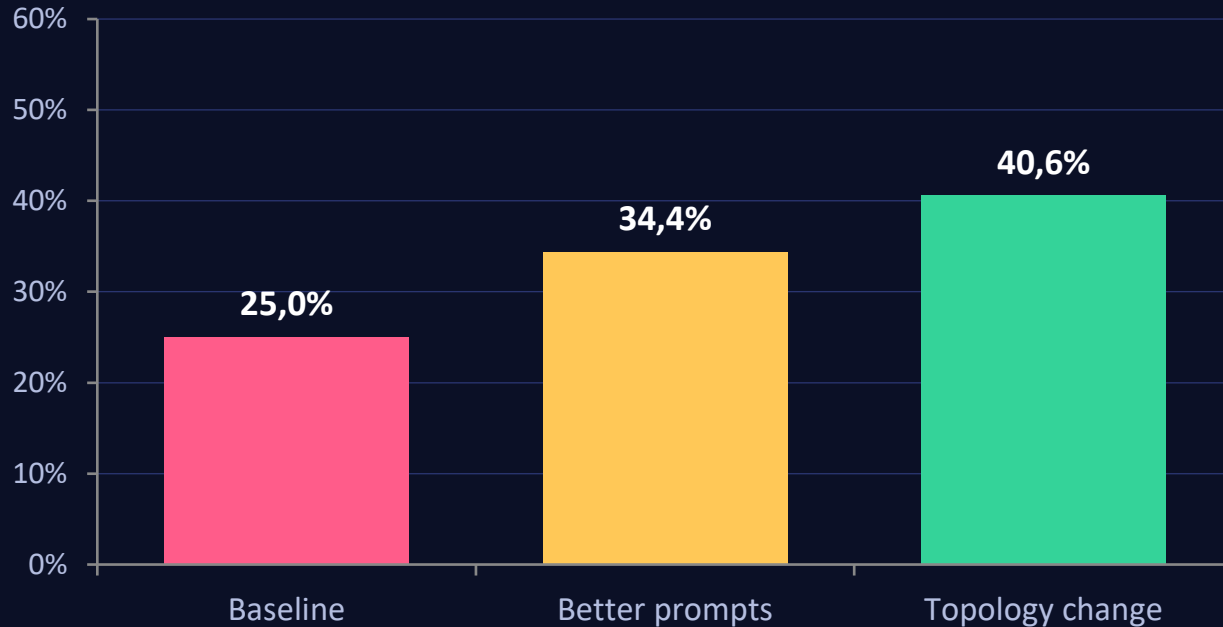
High performers mitigate security risks more often (2026).

− Negative implications

Explainability is widely cited and little mitigated.

Even tuned agents stay below half

ChatDev, 32 tasks: accuracy (%)



- **Gains:** +9.4 and +15.6 percentage points from design tweaks.
- **Most common failures:** repeating steps (15.7%), reasoning that does not become action (13.2%), not recognizing when the task is done (12.4%).
- **Protocols:** MCP and A2A do not solve coordination failures, which occur even in natural language.

★ Importance

Shows the current ceiling of tactical improvements in multi-agent systems.

+ Positive implications

Design tweaks raised accuracy by 9.4 and 15.6 points.

− Negative implications

40.6% is still low for critical tasks without supervision.

Decision rule and adoption figure (Gartner)

Decisions need to be made

→ **AI agents**

Routine workflow

→ **Automation**

Simple information retrieval

→ **Assistants**

17%

of organizations have
already deployed
agents (Gartner,
2026)

60%+

expect to deploy
within two years

Importance

Gives a practical criterion for choosing the right solution.

Positive implications

Well-chosen cases deliver value with less risk.

Negative implications

Fully autonomous agents still do not fit most uses.



What holds

Claim	Verdict	Basis
Gartner predicted canceling over 40% of agentic projects by the end of 2027.	Holds	Official press release (Gartner, 2025a)
The pilot-to-production gap, cost and risk explain much of the problem.	Holds	McKinsey (2025; 2026); RAND (2024); Xu et al.; Yao et al.
"Agent washing" exists in the market.	Holds	Gartner (2025a)
Organizational causes (problem, data, infrastructure) dominate AI failures.	Holds, with a biased-sample caveat	RAND (2024)
Only 17% of organizations have deployed AI agents.	Holds	2026 CIO survey, cited by Gartner (2026)



Importance

Separates what is backed by a source from what is noise.



Positive implications

Five claims have a solid, checked basis.



Negative implications

Even the solid ones have sample or scope limits.

What holds in part

Claim	Verdict	Basis
Design and coordination failures, and not only the model, contribute to agent failures.	Holds in part	Cemri et al.: this is the authors' thesis (conjecture); focuses on design-addressable failures; multi-agent only
"Agentic AI does not work."	Holds in part	Gartner projects growth through 2028; McKinsey (2026): 40% of large companies scale agents
"Over 80% of AI projects fail" (RAND).	Holds in part	RAND writes "by some estimates" and cites Fortune (2022); not its own result
"95% of generative AI pilots deliver no return" (MIT).	Holds in part	Preliminary report (153 respondents, 52 interviews, 300 cases); original not checked

★ Importance

Shows popular numbers that hold only within a given scope.

+ Positive implications

Each scope is useful when cited with context.

- Negative implications

Cited without context, they lead to wrong conclusions.



What does not hold or was not verified

Claim	Verdict	Basis
The 40% is a measured result.	Does not hold	It is an analyst projection; the method is not disclosed in the press release
The 40% stems from the poll of 3,412 attendees.	Does not hold	The press release does not state that link; the poll measures investment intent
"RAND analyzed 2,400 initiatives."	Does not hold	The report is based on 65 interviews; the claim appears on blogs, not at RAND
2026 Deloitte and Forrester figures on agents in production.	Not verified	They appeared on aggregator sites; primary sources were not opened



Importance

Clears up common errors circulating about the numbers.



Positive implications

Avoids decisions based on nonexistent metrics.



Negative implications

Some claims are still not verifiable at a primary source.

Do not add 40%, 80% and 95%

40%

Gartner (2025): forecast of cancellation of agentic projects by 2027

80%

RAND (2024): third-party estimate (Fortune, 2022) about AI projects in general

95%

MIT NANDA (2025): generative AI pilots with no measurable return; preliminary report

★ Importance

Unit, period and definition of failure differ for each number.

+ Positive implications

Citing each number in its own context makes the debate more honest.

- Negative implications

Adding or comparing them leads to unsupported conclusions.

Comparative table of sources

Source	What it measures	Basis	Definition of failure	Period
Gartner (2025a)	Forecast of cancellation of agentic AI projects	Paid report; poll of 3,412 (investment)	"Canceled", with no public definition	Through end of 2027
RAND (2024)	Failure causes as perceived	65 interviews; excludes pretrained-LLM-only	Failure as perceived by the organization	Aug–Dec 2023
MIT NANDA (2025)	Return on generative AI pilots	153 leaders, 52 interviews, 300 cases	No measurable impact on the bottom line	Jan–Jun 2025
McKinsey (2025; 2026)	Self-reported use, scale and impact	1,993 (2025) and 1,719 (2026) respondents	Does not measure failure; measures scale and EBIT	2025 and 2026
Cemri et al. (2025)	Multi-agent task failures	1,642 runs of 7 systems	Not meeting the task objective	2024–25 models

 Importance

Shows why the numbers cannot be added.

 Positive implications

Only Gartner addresses cancellation; the others converge on difficulty.

 Negative implications

The defensible reading is qualitative, not numerical confirmation of the 40%.

Strategy and value: start with the right case

Measure	How to apply	Evidence	Suggested indicator (author)
Choose cases with clear value	Pursue agents only where there is defined value or ROI, and redesign the workflow from the ground up.	Gartner (2025a)	Expected ROI defined before the pilot
Match the tool to the type of task	Agents for decisions, automation for routine, assistants for simple retrieval.	Gartner (2025a)	Rationale for the solution type per use case
Focus on the problem, not the technology	Align technical and business teams on the goal and metrics; commit the team for at least a year.	RAND (2024)	Metric approved by the business owner
Redesign workflows	Redesign the process around AI, with committed leadership; about 3/4 of high performers do this.	McKinsey (2026)	Documented redesigned workflow

★ Importance

Clear value from the start is what separates projects that advance from those canceled.

+ Positive implications

Reduces cost and increases the chance of measurable return.

– Negative implications

Requires discipline to say "no" to cases without value.

Engineering and reliability

Measure	How to apply	Evidence	Suggested indicator (author)
Measure reliability, not just the average	Evaluate the agent over repeated runs (pass^k) and on long tasks, beyond average accuracy.	Yao et al.; Xu et al.	pass^k in real scenarios
Treat failure as an architecture problem	Define roles, cross-agent checks and clear termination criteria.	Cemri et al. (2025)	Failure rate by mode (MAST)
Define human validation	Establish when and how model outputs require human validation.	McKinsey (2025)	% of outputs validated by risk level
Invest in data and infrastructure	Data governance, pipelines and deployment before scaling.	RAND (2024)	Data quality measured per source

★ Importance

Turns reliability into a measurable requirement, not a promise.

+ Positive implications

Design failures are fixable and verifiable.

− Negative implications

Requires investment in testing, data and supervision before scaling.

Cost, security and vendors

Measure	How to apply	Evidence	Suggested indicator (author)
Control cost from the pilot	Measure total cost, including tokens, against the cost of the work replaced.	McKinsey (2026)	Cost per completed task
Manage security risks and unauthorized actions	Mitigate technical vulnerabilities and unintended actions; high performers do this more.	McKinsey (2026)	Incidents per month
Require vendors to prove autonomy	Ask for a demonstration on real data, to avoid agent washing.	Author's inference (Gartner, 2025a)	Demonstration in the customer's environment

For the GRC and information security track: human validation, risk and data governance are the items of greatest relevance. In the author's interpretation, designing controls together with the use case avoids discovering their cost only after the pilot.

★ Importance

Addresses one of the three reasons Gartner cites: inadequate risk controls.

+ Positive implications

Cost and risk measured early avoid surprises at scale.

− Negative implications

Controls and FinOps add effort and cost to the project.

Plausible in direction, not verifiable in number

The forecast

Gartner's 40% is an analyst projection, plausible in direction and not verifiable in number.

The evidence

Converges on three points: the distance between pilot and production, organizational and data causes, fragile reliability.

The 2026 picture

Growing adoption with uneven returns: 40% of large companies scale agents and 37% see some EBIT.

The path

Clear value, the right tool, measure pass^k, human validation, data, and cost and vendors under control.

Only time confirms the real rate: track cancellations in 2027, with "canceled" defined.

What this study cannot claim

- **Gartner:** the full paid report was not accessed.
- **MIT NANDA:** the original was not checked directly.
- **Benchmarks:** use 2024 models.
- **2025 Hype Cycle:** positions come from secondary coverage.
- **McKinsey:** self-reported surveys.
- **AI winters:** history relies on a secondary source.



Next step: track the real cancellation rate in 2027, with the definition of "canceled" made explicit.



News sources: Gartner

Reference (abridged ABNT style)	Status
GARTNER. Gartner predicts over 40% of agentic AI projects will be canceled by end of 2027. Sydney, Jun 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	Original read
GARTNER. Gartner predicts 30% of generative AI projects will be abandoned after proof of concept by end of 2025. Sydney, Jul 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	Official/abstract
GARTNER. Gartner identifies the top 10 strategic technology trends for 2025. Orlando, Oct 21, 2024. https://www.gartner.com/en/newsroom/press-releases/2024-10-21-gartner-identifies-the-top-10-strategic-technology-trends-for-2025	Official/abstract
GARTNER. Hype cycle for artificial intelligence, 2025. Jun 11, 2025 (public abstract). https://gartner.com/en/documents/6579402	Official/abstract
GARTNER. What the 2026 Hype Cycle for Agentic AI reveals. Rajesh Kandaswamy, Apr 15, 2026. https://www.gartner.com/en/articles/hype-cycle-for-agentic-ai	Original read

News sources: reports and surveys

Reference (abridged ABNT style)	Status
RYSEFF, J.; DE BRUHL, B. F.; NEWBERRY, S. J. The root causes of failure for artificial intelligence projects and how they can succeed. Santa Monica: RAND, 2024. 20 p. (RR-A2680-1). DOI 10.7249/RRA2680-1. https://www.rand.org/pubs/research_reports/RRA2680-1.html	Original read
SINGLA, A. et al. The state of AI in 2025: agents, innovation, and transformation. McKinsey & Company, Nov 5, 2025. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-2025	Original read
TINKOFF, D.; VAN DER VEKEN, L.; CHUI, M.; BALAKRISHNAN, T. The state of AI in 2026: on the road to ROI. McKinsey & Company, Aug 25, 2026. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai	Original read
CHALLAPALLY, A. et al. The GenAI Divide: state of AI in business 2025. MIT NANDA, Jul 2025 (preliminary). Third-party copy: https://webberwentzel.com/News/Documents/2025/MIT1757411281972.pdf	Secondary
KAHN, J. Want your company's A.I. project to succeed? Fortune, 26 jul. 2022. Cited in Ryseff et al. (2024); not consulted directly.	Secondary

Academic studies

Reference (abridged ABNT style)	Status
CEMRI, M. et al. Why do multi-agent LLM systems fail? arXiv:2503.13657, 2025 (v3). https://arxiv.org/abs/2503.13657	Original read
XU, F. F. et al. TheAgentCompany: benchmarking LLM agents on consequential real world tasks. arXiv:2412.14161, 2024. https://arxiv.org/abs/2412.14161	Official/abstract
YAO, S.; SHINN, N.; RAZAVI, P.; NARASIMHAN, K. τ -bench: a benchmark for tool-agent-user interaction in real-world domains. ICLR 2025. arXiv:2406.12045. https://arxiv.org/abs/2406.12045	Official/abstract
YAO, S. et al. ReAct: synergizing reasoning and acting in language models. arXiv:2210.03629, Oct 6, 2022. https://arxiv.org/abs/2210.03629	Official/abstract
PARK, J. S. et al. Generative agents (arXiv:2304.03442, 2023); QIAN, C. et al. ChatDev (2307.07924); HONG, S. et al. MetaGPT (2308.00352); WU, Q. et al. AutoGen (2308.08155). Identifiers as listed in the references of Cemri et al. (2025).	Official/abstract



Sources for the timeline and history

Reference (abridged ABNT style)	Status
OPENAI. Introducing ChatGPT. Nov 30, 2022. https://openai.com/blog/chatgpt (date checked in secondary sources).	Secondary
AUTOGPT. Wikipedia. Released Mar 30, 2023 (confirmed by other sources). https://en.wikipedia.org/wiki/AutoGPT	Secondary
COGNITION unveils Devin: autonomous AI software engineer. Dealroom, Mar 2024 (13.86% on SWE-bench). https://nz.dealroom.co/news/note/cognition-unveils-devin-autonomous-ai-software-engineer	Secondary
Protocols for agentic AI: Google's new A2A joins viral MCP. Virtualization Review, Apr 9, 2025. https://virtualizationreview.com/articles/2025/04/09/protocols-for-agentic-ai-googles-new-a2a-joins-viral-mcp.aspx · O'REILLY Radar (MCP and A2A in 2025–2026): https://www.oreilly.com/radar/?p=19654 · ANTHROPIC. Building effective agents (Dec 2024), cited by Cemri et al.	Secondary
DATA CAMP. AI winter: history and timeline. https://datacamp.com/blog/ai-winter · LIGHTHILL, J. Artificial intelligence: a general survey. Science Research Council, 1973 (not re-consulted).	Secondary