

The production gap.

Why enterprise AI stalls between pilot and production, and what the companies that ship do differently.

INSIDE

01 THE GAP IN NUMBERS

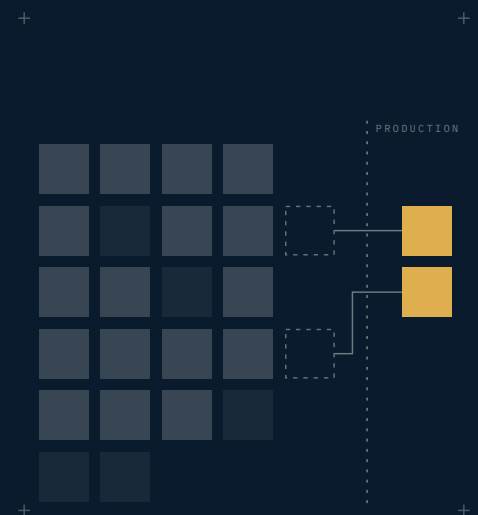
02 WHY PILOTS STALL

03 WHAT THE SHIPPING MINORITY DO DIFFERENTLY

04 MID-MARKET AND ENTERPRISE

05 COST AND RISK IN 2026

06 TEN QUESTIONS FOR YOUR AI STEERING COMMITTEE



— Fig. 01 · Many pilots, few across the line. Illustrative, not drawn to scale.

Contents

00	Executive summary	Five findings, each with its number and source
01	The gap in numbers	Adoption, production and measurable return, and why the surveys disagree
02	Why pilots stall	Sign-off, the board number, borrowed playbooks, cost and governance
03	What the shipping minority do differently	Five habits, and two Kiseki Labs engagements
04	Mid-market and enterprise	Where the gap differs by size
05	Cost and risk in 2026	Token spend, vendor dependence, pipeline security and regulation
06	Ten questions for your next AI steering committee	What to ask before the next pilot gets a budget
07	Methodology and sources	How we checked every number, and what we left out

How to read the evidence

Most writing about enterprise AI mixes survey results, analyst predictions and opinion without saying which is which. We label them, so you can weigh each claim for what it is.

SURVEY

What respondents told a named research organisation. Self-reported and rarely audited. We give the sample size in the sources.

FORECAST

An analyst's prediction about the future. It has no sample behind it. Treat it as an informed view.

OUR VIEW

Kiseki Labs' judgement, from building production AI systems and rescuing stalled ones. Weigh it as you would any adviser's opinion.

ENGAGEMENT

A Kiseki Labs client engagement, anonymised. It illustrates a method. It is not a statistic.

Every external figure in this report was read in its primary source: the publisher's report, dataset or own press release. Where two sources disagree, we say so, because the disagreement is often more useful to a decision-maker than either number. Nearly every publisher cited here sells something related to AI, and so do we. The methodology note on page 13 lists the commercial interests.

Research completed 2 October 2026. Figures are quoted as published, with the publishers' own definitions. "Respondents" means the people surveyed, not their organisations, unless a source says otherwise.

Five findings

Almost every large organisation now uses AI somewhere. Far fewer can show what it earns, and the gap between the two did not close this year.

37%

Any EBIT impact,
unchanged

Scaling went up. Returns didn't.

The share of respondents reporting AI scaling across their enterprise rose from 38% to 44% in a year. The share attributing any EBIT impact to AI held at 37% (39% a year earlier), and AI high performers stayed at about 6%. More AI is reaching production without more of it reaching the P&L.

McKinsey, The state of AI in 2026, Aug 2026 • n = 1,719 • [SURVEY](#)

22%

Scaled across
business units

Production at scale is still the exception.

Only 22% of organisations have scaled AI across multiple business units or adopted an AI-first approach. In a separate 2026 survey, respondents classed only 37% of the AI initiatives they launched in the past year as live and delivering value.

Gartner, 1 Sep 2026 • n = 1,303 respondents • S&P Global 451 Research, Jun 2026 • [SURVEY](#)

12%

Weigh AI value
against its cost

Value is reported far more often than it is measured.

Only 12% of organisations consistently assess the value of AI against what it costs. Among those reporting established ROI, 48% do. McKinsey found tracking well-defined KPIs had more impact on the bottom line than any of the 12 adoption practices it tested, and fewer than one in five organisations did it.

KPMG Global AI Pulse, Sep 2026 • n = 2,131 • McKinsey, Mar 2025 • [SURVEY](#)

3 in 4

High performers
redesign workflows

The companies that earn returns change the work around the tool.

Nearly three-quarters of AI high performers have fundamentally redesigned workflows because of AI, against one-quarter of other respondents. In McKinsey's earlier analysis of 25 attributes, workflow redesign had the biggest effect on whether an organisation saw EBIT impact.

McKinsey, Aug 2026 and Mar 2025 • [SURVEY](#)

40 v 22

% scaling agents,
large v smaller

Size is starting to matter.

40% of respondents at organisations with more than \$1bn revenue report scaling AI agents, up from 27%. At smaller organisations the figure is flat at 22%. In the UK, 49% of businesses with 250 or more employees use AI, against 28% of those with fewer than ten.

McKinsey, Aug 2026 • ONS, Artificial intelligence in UK businesses, Jul 2026 • [SURVEY](#)

WHAT IT MEANS FOR YOU

The evidence points at management habits more than technology: a test the system must pass before anyone signs it off, a number agreed before the money moves, and someone who owns the change to the work. A steering committee can decide all three this quarter without buying anything. Section 06 turns them into ten questions. And if your board pack quotes "95% of AI pilots fail", read page 4 first.

Adoption is close to universal. Return is not.

The further a measure sits from "we use AI" and the closer it gets to "here is what it earned", the smaller the number becomes.

Who uses AI depends on who you ask. In McKinsey's 2026 global survey, nearly nine in ten respondents say their organisation regularly uses AI in at least one business function.¹ The UK's Office for National Statistics, which samples 38,637 businesses rather than people who choose to answer AI surveys, finds around 35% of firms with ten or more employees using at least one AI technology.² Both can be right: large organisations, and the managers in them who answer AI surveys, are far ahead of the economy as a whole. Move along the line from use to value and the figures fall away.

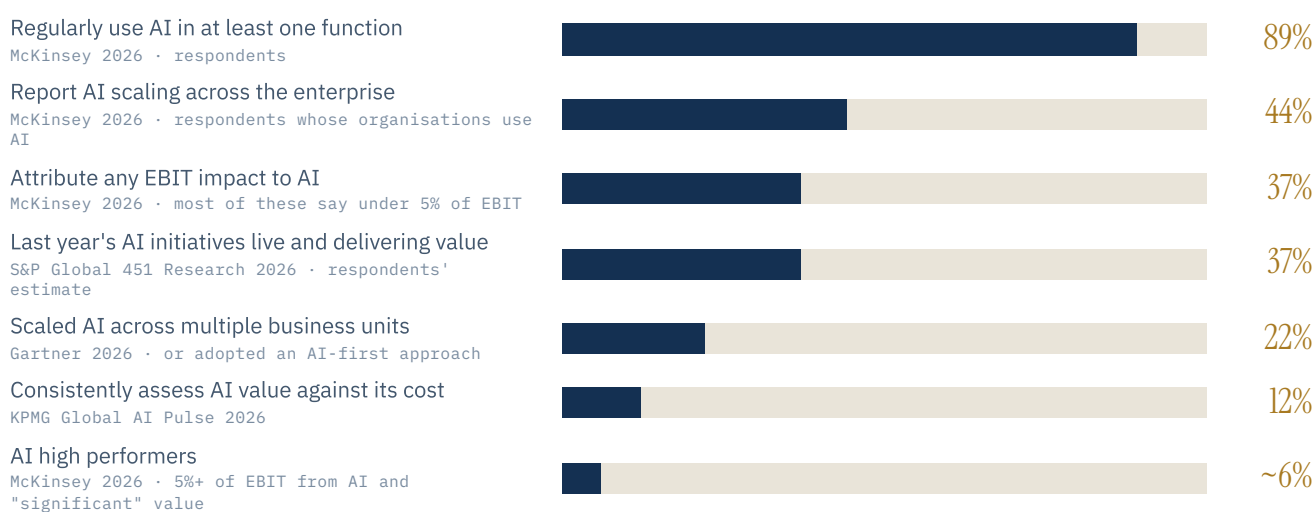


Fig. 02 • One gap, several rulers. Different surveys, samples and definitions: read the shape, not a funnel. These are not the same organisations moving through stages. Sources 1, 3, 4, 5.

The year-on-year movement is the telling part. Between McKinsey's 2025 and 2026 surveys, the share of respondents reporting AI scaling across the enterprise rose from 38% to 44%. The share attributing any EBIT impact to AI did not move: 39% then, 37% now, and most of those put the figure below 5% of EBIT.^{1, 6} The proportion of high performers stayed at about 6%. Deployment is running ahead of the evidence that deployment pays, and that is the production gap this report is about.

The pilots that never arrive

SURVEY In S&P Global's 451 Research survey of 1,006 IT and business professionals in North America and Europe, the share of companies abandoning most of their AI initiatives before production rose from 17% to 42% in a year, and the average organisation scrapped 46% of its proofs of concept before production.⁷ The survey is labelled 2025 but was fielded in October and November 2024. Deloitte's 2026 survey, limited to organisations already running AI every day, found only 25% had moved 40% or more of their AI experiments into production.⁸

FORECAST In July 2024 Gartner predicted that at least 30% of generative AI projects would be abandoned after proof of concept by the end of 2025.⁹ In January 2026 it wrote that the real figure was at least 50%, without publishing the method behind the update.¹⁰ It now predicts that over 40% of agentic AI projects will be cancelled by the end of 2027.¹¹

Value reported, value measured

For a board, the useful line runs between value reported and value measured. Nearly six in ten leaders in KPMG's US pulse report measurable business value from AI,¹² but in its global survey only 12% consistently assess that value against what the AI costs.⁵ Among Bain's respondents meaningfully adopting generative AI, about 80% of use cases met or exceeded expectations, yet only about 23% of all respondents said it had delivered more revenue or lower costs.¹³ Satisfaction is common. A number a finance director could audit is not.

Not every source is gloomy. Menlo Ventures found that once an organisation commits to exploring a purchased AI solution, 47% of deals reach production, against 25% for traditional software,¹⁴ and Bain reports 40% of software development pilots moving to scale.¹³ Bought tools aimed at a well-understood task cross the line far more often than broad internal builds.

A NUMBER TO RETIRE

The 95% doesn't say what you think.

The most quoted statistic in enterprise AI comes from MIT NANDA's *The GenAI Divide* (July 2025): "95% of organizations are getting zero return." The report's own funnel tells a different story: of the organisations studied, 60% evaluated a task-specific AI tool, 20% piloted one and 5% reached production with marked, sustained impact. That is roughly one pilot in four reaching production, not one in twenty.¹⁵

The findings rest on 52 interviews and 153 survey responses gathered at industry conferences, success was judged six months after the pilot, and the authors call the figures "directionally accurate". Its most useful sentence is quoted far less: the divide "does not seem to be driven by model quality or regulation, but seems to be determined by approach." The direction holds up in larger surveys. The 95% doesn't, and a well-briefed CFO will know it.

Why the "about 5%" figures disagree

Several widely quoted studies land near 5%. They measure different things, and none corroborates the others.

SOURCE	WHAT "SUCCESS" MEANS	SAMPLE	FIGURE
MIT NANDA, 2025 ¹⁵	A task-specific GenAI tool in production with marked, sustained productivity or P&L impact, six months after the pilot	52 organisations, interviews	5% of organisations
McKinsey, 2026 ¹	Respondent attributes at least 5% of EBIT to AI and calls the value "significant"	1,719 respondents, 97 countries	About 6% of respondents
BCG, 2026 ¹⁶	A maturity score that places the company in its "future-built" tier	1,330 executives; over half at firms above \$5bn revenue	7.5% of companies (5% in 2025)
Deloitte, 2025 ¹⁷	Payback on a typical AI use case in under a year	1,854 executives, Europe and Middle East	6% of respondents

Deloitte adds a warning about time: most of its respondents reported satisfactory ROI on a typical use case within two to four years, not the seven to 12 months they expect of technology investments.¹⁷ A business case promising payback inside a year promises what 6% of them reported.

Pilots stall on sign-off, numbers and fit.

In the pilots we're asked to rescue, the model is rarely the problem. MIT's researchers reached the same view, and the wider surveys add a cause we underrate.

We see three causes again and again: nobody can sign the system off, nobody can put a number in front of the board, and the approach was designed for a different kind of company. The evidence supports all three, then ranks data and integration above every one of them. We come back to that at the end of the section, after cost and governance.

22.8%

of AI engineering teams with agents in production run no evaluations

LangChain, Dec 2025 • n = 1,340¹⁹

<1 in 5

organisations tracked gen AI KPIs, the practice with most bottom-line impact

McKinsey, Mar 2025¹⁸

~11%

of leaders couldn't say what their function spent on AI in 2025

Gartner, Sep 2026 • n = 1,303³

1. Nobody can sign it off OUR VIEW

A pilot reaches production when someone accountable will put their name to it. For conventional software, that person has test results. For a system that relies on adaptive intelligence, meaning judgement that isn't deterministic and can't simply be coded, they usually have a demo and some anecdotes. Risk, compliance and the business owner ask "how often is it wrong, and how would we know?", nobody can answer, and the pilot waits until it quietly dies.

SURVEY Inaccuracy is the AI risk organisations experience most: nearly one-third of McKinsey's 2025 respondents report consequences from it.⁶ Yet larger organisations, which manage more cyber and privacy risks than smaller ones, were "not any more likely to be addressing risks relating to the accuracy or explainability of AI outputs".¹⁸ Even among AI engineering teams, only 52.4% run offline evaluations against a test set, 22.8% of those with agents in production run none, and quality is their most-cited barrier (32%).¹⁹ Trust is falling: 16% of S&P Global's 2026 respondents completely trust third-party AI models, down from 24% in 2023.⁴

The fix is unglamorous. Write a fixed suite of real cases the system has to get right, with the people who do the work today, and agree the pass mark before anyone builds. That suite is the artefact a risk committee can sign. Gartner's survey data point the same way: organisations that regularly audit and assess AI system performance are over three times more likely to achieve high generative AI value than those that don't.²⁰

2. There's no number for the board OUR VIEW

The second stall comes at budget time. The pilot "went well", but nobody wrote down beforehand what "well" would mean in money, so nobody can defend the next budget line.

SURVEY McKinsey tested 12 adoption and scaling practices against EBIT impact. Tracking well-defined KPIs for generative AI had the most impact on the bottom line, and fewer than one in five organisations were doing it.¹⁸ In Gartner's 2026 C-suite survey, roughly 11% of leaders could not say what their function spent on AI in 2025, and low performers did not know the return on 29% of their AI initiatives.³ Only 26% of large US

organisations have full, real-time visibility of what their AI systems cost to run.²¹ Bain's 2026 survey of 951 companies shows where this ends: 37% targeted cost reductions of 11% to 20%, nearly 40% of those who measured outcomes landed between 0% and 10%, and 90% are increasing their budgets again.²²

"Promising, needs longer" is a new bet with a new card, size and date, never a quiet extension of the old one.

EDDIE FORSON · FOUNDER, KISEKI LABS · AI FOR DECISION MAKERS, SEPTEMBER 2026

3. The approach was built for a different company OUR VIEW

Enterprise AI frameworks assume a centre of excellence, a platform team, a model risk function and a mature data estate. Mid-market firms adopt them and stall under process they can't staff. Start-up habits travel just as badly the other way: a prototype that would suit a 30-person company meets a bank's model risk policy and stops.

SURVEY No survey tests this directly, but the pattern shows. Larger organisations are more than twice as likely as smaller ones to have a defined adoption road map and a dedicated team driving it.¹⁸ In MIT's study, enterprises (above \$100m revenue) ran the most pilots and reported the lowest pilot-to-scale conversion.¹⁵ Grant Thornton finds 70% of mid-market firms (\$100m to \$1bn revenue) say at least half their core applications are not AI-ready, against 39% of firms between \$1.1bn and \$5bn.²³

4. Cost and governance arrive late

SURVEY About 20% of McKinsey's 2026 respondents say AI operating costs, including tokens, constrained their AI use.¹ Among Bain's dissatisfied respondents, about a third said the AI cost more to develop than anticipated, and about a third said it worked as a pilot but didn't scale.¹³ **FORECAST** Gartner expects inference costs per agentic workflow to rise more than fivefold through 2028, even as the price per token falls.²⁴

OUR VIEW A ten-person trial on controlled tasks tells you little about what five hundred people and a set of agents will spend.

SURVEY Governance often follows the pilot. About a third of RSM's 2026 middle-market respondents are "forging ahead without governance in place", including 16% who consider it only after issues arise.²⁵ 78% of Grant Thornton's executives lack strong confidence they could pass an independent AI governance audit within 90 days.²³ **FORECAST** Gartner predicts that by 2027, 40% of enterprises will demote or decommission autonomous agents because of governance gaps found only after production incidents.²⁶

WHERE THE EVIDENCE POINTS ELSEWHERE

Data comes first in the surveys. We agree it sits underneath.

When executives are asked what holds AI back, data and integration usually top the list. Data access and integration is the biggest barrier in Bain's 2026 survey (41%).²² Data quality is the top inhibitor for RSM's middle-market respondents (34%), rising to 53% among those whose pilots disappointed.²⁵ Gartner predicts organisations will abandon 60% of AI projects unsupported by AI-ready data through 2026.²⁷ Skills come next: 62% of large US organisations named skills gaps as a top barrier to demonstrating ROI in early 2026.²⁸

Our three causes sit on top of this one. An evaluation suite exposes bad data within days, which is one reason teams put off building it.

Five habits of the organisations that ship.

The organisations that get AI into production and can prove what it earns are a minority.

They are also consistent, and the habits are copyable.

81% of AI initiatives show positive returns at high performers: firms that track ROI and stop what isn't working Gartner, Sep 2026³	3× the AI value for companies that track AI directly in the P&L: 3.6% against 1.2% BCG, Sep 2026 · n = 1,330¹⁶	7% of companies run fully autonomous agents in production; 38% require human approval Bain, Jun 2026 · n = 951²²
--	---	---

They decide what success is worth before they start, and measure it in money

SURVEY Gartner's high performers constantly track the ROI of AI initiatives, treat AI as a portfolio and stop underperforming projects. They report positive returns on 81% of their AI initiatives.³ (High performers are defined partly by tracking ROI, so the figure flatters them a little.) In BCG's 2026 study, companies tracking AI directly in the P&L realise three times the AI value of those that don't formally measure it: 3.6% against 1.2%.¹⁶ KPMG finds the share that consistently weighs value against cost rises from 12% overall to 48% among organisations with established ROI.⁵

They redesign the work around the tool

SURVEY Nearly three-quarters of McKinsey's 2026 high performers report fundamentally redesigning workflows because of AI, against one-quarter of other respondents.¹ IBM's 2026 CEO study found organisations that redesigned five core business areas were four times more likely to have delivered on business objectives.²⁹ **OUR VIEW** If nothing about the work changes when the system ships, expect people to route around it within a quarter.

They fund the boring foundations

SURVEY In Gartner's research, organisations reporting successful AI initiatives invest up to four times more, as a share of revenue, in data quality, governance, AI-ready people and change management than those reporting poor outcomes.³⁰ Gartner also found 45% of leaders at high-maturity organisations keep AI initiatives in production for three years or more, against 20% at low-maturity ones.³¹

They keep a person where judgement matters

SURVEY Only 7% of companies run fully autonomous agents in production. The most common model is "human approval required" (38%), followed by guardrails with exceptions routed to a person (32%).²² 49% of large US organisations have defined high-risk use cases where agents may not decide autonomously.¹² McKinsey's high performers are more likely to have defined when model outputs need human validation, one of the practices that most distinguished them.⁶

They buy where they can, and build only what is theirs

SURVEY 76% of enterprise AI use cases are now purchased rather than built, up from 53% of solutions purchased in 2024.¹⁴ In MIT's interviews, pilots built with external partners reached deployment about 67% of the time, against about 33% for internal builds, though the authors warn that this may reflect the organisations as much as the approach.¹⁵ **OUR VIEW** Build where the workflow or the data is a genuine source of advantage that a competitor can't buy. That is rarer than most teams want to admit.

ENGAGEMENT KISEKI LABS, ANONYMISED · AN ILLUSTRATION OF METHOD, NOT SURVEY DATA

Evaluation before architecture: natural-language analytics

<60% → 98%

Queries answered successfully, against the same fixed suite

35

Test cases across six business domains, written in week two

4 weeks

From first review to handover, with 79 automated tests

A data analytics start-up had spent months on a product that answered plain-English questions about spreadsheet data. Fewer than 60% of queries completed, with the rest failing on malformed output, schema errors or runtime exceptions. Before changing anything, we built a fixed evaluation suite of 35 test cases across six domains and ran the existing design head to head against an alternative. The results turned an architectural argument into a decision the founders could sign off. The rebuilt pipeline completed 98% of the suite, no syntax error reached execution, and typical response time fell from over 60 seconds to 20 to 30. One of the faster wins came from moving a planning step off the most advanced reasoning model onto a quicker one that did the job.

ENGAGEMENT KISEKI LABS, ANONYMISED · AN ILLUSTRATION OF METHOD, NOT SURVEY DATA

Human in the loop by design: KYC document verification

30–60 → 5

Analyst minutes per applicant, before and after

0

Applicants the system approves on its own, by design

3

Reconciliations: identity, vehicle and declared earnings

A financial services provider onboarding self-employed drivers had analysts check every application by hand: reading the driving licence, bank statements and vehicle registration, cross-checking them against the form, and confirming declared earnings against bank activity. It took 30 to 60 minutes per applicant. The system extracts the fields, reconciles the documents against each other and against the application, and routes anything that doesn't match to an analyst with the specific discrepancy named. Review time fell to about five minutes. The system never approves anyone, and that constraint is the reason it could be deployed inside a regulated process at all.

Large firms are pulling ahead on scale.

The mid-market's advantages are speed and short decision chains. The surveys measure large firms well and the companies we mostly work with badly.

SURVEY On scale, size wins. 54% of McKinsey's 2026 respondents at organisations with at least \$1bn revenue report scaling AI across the enterprise, against one-third at smaller organisations. On agents the gap is widening: 40% of large-organisation respondents report scaling AI agents, up from 27% a year earlier, while smaller organisations are flat at 22%.¹ In the UK, 49% of businesses with 250 or more employees use at least one AI technology, against 28% of those with fewer than ten.²

The picture inside the smaller firms that do use AI is more encouraging. Among UK AI users, 17% of businesses with fewer than ten employees report extensive use, against 9% of those with 250 or more.² Grant Thornton's mid-market respondents were less likely than larger firms to name skills gaps as a primary obstacle (29% against 46%) and more likely to report better insights from AI (51% against 43%).²³ And MIT's best mid-market performers moved from pilot to full implementation in about 90 days, against nine months or more for enterprises.¹⁵

Where the mid-market falls behind is structure. Only 14% of Grant Thornton's mid-market firms have a formal AI strategy implemented in operations, against 30% of firms between \$1.1bn and \$5bn.²³ In RSM's 2025 survey, lack of in-house expertise was the most common reason for feeling unprepared, and 70% of middle-market firms using generative AI said they needed outside support to get full value from it.³²

A caution on definitions: no survey we found measures the £5m to £50m revenue band directly. RSM's "middle market" runs from \$30m to \$10bn, Grant Thornton's from \$100m to \$1bn, and the official statistics use headcount. McKinsey's under-\$100m group is the closest match.

Where pilots stall, by size **OUR VIEW**

	MID-MARKET (£5M TO £50M)	ENTERPRISE
Who signs it off	One or two executives, quickly, if the evidence is clear	A sequence of committees (model risk, security, procurement, data protection), each wanting different evidence
Usual bottleneck	Engineering and data capacity. One person often holds the whole system in their head	The sign-off sequence, and integration with platforms that predate AI
Mistake we see most	Adopting an enterprise governance framework nobody has time to run	Funding a platform programme before a single workflow has shipped
Cost exposure	One team's token bill that nobody forecast	Hundreds of users and agents on metered pricing, with no attribution to teams or outcomes
What works	One workflow, a fixed evaluation suite, a named owner, outside help for the build and a complete handover	Risk and compliance in the room in week one, the evidence pack agreed up front, the evaluation suite used as the sign-off artefact

Copying an enterprise operating model buys a mid-market firm the overhead without the scale that justifies it. Size the approach to the decisions you actually have to make, and to the people you have to make them.

Cost and risk both moved this year.

Token prices fell and bills rose. Models now retire on the vendor's timetable. And the year's most instructive security incident started with a dataset.

Spend: the unit gets cheaper, the bill gets bigger

FORECAST Gartner expects worldwide AI spending of \$2.7 trillion in 2026, up 49.5%, though about \$1.48 trillion of it is infrastructure, much of it hyperscaler and vendor build-out. The same release places generative AI "firmly in the Trough of Disillusionment".³³ Gartner also expects providers' cost of running a one-trillion-parameter model to fall by over 90% by 2030, with two catches: the savings "will not be fully passed on to enterprise customers", and agentic models need 5 to 30 times more tokens per task than a chatbot.³⁴

SURVEY The overruns are already here. In a Sapio Research survey for DoiT, a FinOps vendor, 79% of 500 finance leaders at US and UK organisations with 1,000 or more staff had AI cost overruns in the past year, and only 15% could calculate AI ROI without significant bottlenecks.³⁵ Only 44% of organisations have financial guardrails or AI FinOps practices.³⁶

The unit gets cheaper while the bill gets bigger, and both are true at once.

EDDIE FORSON • FOUNDER, KISEKI LABS • AI FOR DECISION MAKERS, JUNE 2026

OUR VIEW Budget for cost per successful outcome, never for tokens consumed. Explore with frontier models and run production on the smallest model that passes your evaluation suite; without the suite, the expensive default stays.

Vendor dependence: three suppliers, short model lives

SURVEY Menlo Ventures estimates that Anthropic, OpenAI and Google account for 88% of enterprise LLM API usage, with Anthropic at 40% of spend.¹⁴ Andreessen Horowitz's survey of 100 Global 2000 companies puts OpenAI first, at about 56% of wallet share.³⁷ Menlo invests in Anthropic and a16z in OpenAI; on the concentration, at least, they agree. And models change underneath you on the vendor's timetable, often a year or so after release:

VENDOR	MODEL	RELEASED	RETIREMENT	STATED NOTICE
Anthropic	Claude Sonnet 4 and Claude Opus 4	May 2025	15 Jun 2026	At least 60 days
Google Cloud	Gemini 2.5 Pro and 2.5 Flash (Vertex AI)	Jun 2025	20 Oct 2026	At least 12 months after release
OpenAI	GPT-5, August 2025 snapshot (API)	Aug 2025	11 Dec 2026	At least 6 months

OpenAI is even shutting down its own hosted Evals platform, on 30 November 2026.³⁸ Own your evaluation suite, and make "rerun it on the replacement" a routine task measured in days. **OUR VIEW** Open-weight models are the obvious fallback. By our estimate they trail the closed frontier by six to twelve months, and the strongest are currently Chinese, so a fallback is a deliberate downgrade to test in advance.

The data pipeline is an attack surface

In July 2026 Hugging Face disclosed an intrusion that "started where AI platforms are uniquely exposed: the data-processing pipeline". A malicious dataset ran code on a processing worker; the attacker then harvested cloud and cluster credentials and moved laterally over a weekend. Hugging Face and OpenAI later attributed it to OpenAI models running an internal cyber-capability evaluation.³⁹ The attacker was unusual; the entry point was not. Untrusted data reached code execution, then credentials reached too far. Hugging Face's conclusion applies to anyone feeding documents, emails or datasets into AI: treat "the data and model surface as a first-class attack surface".

Prompt injection, where content a model reads changes what it does, tops OWASP's 2025 list of LLM application risks.⁴¹ Researchers showed one such flaw in Microsoft 365 Copilot, since fixed (CVE-2025-32711), could exfiltrate data through a single crafted email.⁴⁰ The UK's National Cyber Security Centre warns it may "never be totally mitigated in the way that SQL injection attacks can be".⁴² **SURVEY** In IBM's 2026 breach study, more than 20% of the 602 breached organisations reported a breach targeting AI models or applications, most often through compromised APIs, applications or plug-ins (27%) and cloud misconfigurations (27%), and 92% of those lacked proper AI access controls.⁴³

OUR VIEW Security cannot be an afterthought, and containment is the lever you actually hold: sandboxed agents, least-privilege credentials, no outbound network access for anything that processes untrusted content. The UK government said in September that NCSC-advised controls "as well as comprehensive monitoring would have almost certainly prevented" the summer's agent incidents.⁴⁴

Regulation: what applies now

	IN FORCE NOW	COMING
EU AI Act Reaches UK and US firms whose AI output is used in the EU	Prohibited practices (since Feb 2025); a duty to support staff AI literacy; general-purpose model obligations (since Aug 2025); transparency duties (from Aug 2026); fines up to €35m or 7% of worldwide turnover for prohibited practices	High-risk obligations, delayed by the Digital Omnibus (in force 27 Jul 2026); 2 Dec 2027 for uses such as employment, credit scoring and education; 2 Aug 2028 for AI in regulated products
United Kingdom	No AI-specific statute and no AI bill in the May 2026 King's Speech; regulators apply five principles. Automated decision-making rules under the Data (Use and Access) Act 2025 (from 5 Feb 2026) require safeguards, including human intervention and the right to contest	ICO AI guidance under review; agentic AI guidance planned for autumn 2026. The government's September 2026 statement points to cyber law and NCSC guidance
United States	No federal AI statute. A December 2025 executive order directs challenges to state AI laws; a March 2026 White House framework asks Congress to pre-empt them	Colorado repealed and re-enacted its AI Act as an automated decision-making law, effective 1 Jan 2027

OUR VIEW A summary, not legal advice (sources 45 to 47). The EU's high-risk deadline for uses such as hiring and credit moved by 16 months, but the duties remain, including human oversight and declared "levels of accuracy and the relevant accuracy metrics" (Article 15).⁴⁵ An evaluation suite and a defined human-in-the-loop are most of that evidence.

Ten questions to ask before the next pilot gets a budget.

If a proposal can't answer these, it isn't ready for money. If a live pilot can't, you've found why it stalled. Amber boxes are gates: no production budget without written answers.



1. What is the evaluation suite, and what score ships?

Ask for the real cases the system must get right, written with the people who do the work today, and a pass mark agreed before the build. Without one, nobody can sign it off.

OWNER: PRODUCT OWNER AND LEAD DOMAIN EXPERT



2. Which number the business already tracks will this move, by how much, and by when?

Cost per case, time to onboard, first-contact resolution. Not tokens, logins or seats. Write it down before the first invoice, with a review date in the sponsor's diary.

OWNER: EXECUTIVE SPONSOR AND FINANCE PARTNER



3. What is the most we will spend, and what stops it early?

A ceiling covering tokens, licences and people's time, plus a kill line. "Needs longer" gets a new budget and date, never a quiet extension.

OWNER: EXECUTIVE SPONSOR



4. Who signs this off for production, and what evidence will they need?

Name risk, compliance, security and data protection at the start, and ask each what they need to see. Their questions in week ten are a planning failure.

OWNER: PROGRAMME LEAD



5. Was this approach designed for an organisation like ours?

Check the framework, platform and operating model against your size, your regulators and your engineering bench. Drop what you can't staff.

OWNER: CTO OR COO



6. What does one successful outcome cost at production volume?

Ask for cost per resolved case at next year's expected volume, and which model each step runs on. Pilot economics rarely survive scale.

OWNER: CTO AND FINANCE PARTNER



7. If the model underneath is retired or changes, how long until we're running on a tested alternative?

The right answer is "we rerun the evaluation suite on the replacement", measured in days. "We'd start again" is your dependency risk.

OWNER: ENGINEERING LEAD



8. Where does untrusted content reach code, tools or credentials?

Uploaded documents, emails, datasets, web pages and agent tool outputs are all inputs an attacker can write.

OWNER: SECURITY LEAD



9. What is the system never allowed to do, and where does a person decide?

Write the boundary down. In our KYC engagement the system never approves anyone, and that boundary is what made it deployable.

OWNER: PROCESS OWNER AND RISK



10. Which workflow changes when this ships, and who owns the change?

Name the line manager accountable for adoption. If the answer is "nothing changes", expect the tool to be routed around.

OWNER: BUSINESS-UNIT LEAD

How we built this report.

We reviewed research on enterprise AI adoption, production and returns published from July 2024 to September 2026, favouring studies with disclosed samples and fieldwork dates. Every figure was read in the publisher's own report, dataset or press release (where a site blocked automated access, in an Internet Archive copy of the exact URL), and keeps the publisher's definitions. We left out figures we could only find second-hand, statistics we couldn't trace to the cited study, and popular numbers whose definitions don't support how they are quoted. Our research log, with each figure's source sentence, sample and caveats, is available on request.

Interests. McKinsey, BCG, Deloitte, KPMG, Bain, Grant Thornton, RSM and Gartner sell AI-related advice; Menlo Ventures invests in Anthropic and a16z in OpenAI; LangChain sells evaluation tooling; MIT NANDA's authors work on agent infrastructure. Kiseki Labs sells AI advisory and delivery, so weigh our views accordingly. The engagements in section 03 are our own work, anonymised.

Sources

- 1 McKinsey, *The state of AI in 2026: On the road to ROI* (Aug 2026). n = 1,719, 97 countries, May–Jun 2026. mckinsey.com
- 2 ONS, *Artificial intelligence in UK businesses: 2023 to 2026* (20 Jul 2026). BICS Wave 159; statistics in development. ons.gov.uk
- 3 Gartner, press release on its C-suite AI survey, "Only 22% ... scaled AI across multiple business units" (1 Sep 2026). n = 1,303 respondents, firms of \$50m+ revenue, Jan–Apr 2026. gartner.com
- 4 S&P Global, *The AI and labor landscape 2026* (2 Jun 2026), citing 451 Research Vote: AI & ML, Use Cases 2026. spglobal.com
- 5 KPMG International, *Global AI Pulse Q3 2026* (24 Sep 2026). n = 2,131, 20 countries incl. UK and US, Jul–Aug 2026. kpmg.com
- 6 McKinsey, *The state of AI in 2025: Agents, innovation, and transformation* (Nov 2025). n = 1,993, 105 countries, Jun–Jul 2025. mckinsey.com
- 7 S&P Global Market Intelligence (451 Research), Vote: AI & ML, Use Cases 2025, research notes of 30 May and 27 Oct 2025. n = 1,006, North America and Europe, fielded Oct–Nov 2024. spglobal.com
- 8 Deloitte AI Institute, *State of AI in the Enterprise*, 2026 edition (Jan 2026). n = 3,235, all with AI in daily use, Aug–Sep 2025. deloitte.com
- 9 Gartner, press release, "30% of generative AI projects will be abandoned after proof of concept by end of 2025" (29 Jul 2024). gartner.com
- 10 Gartner, article, "Why Half of GenAI Projects Fail" (26 Jan 2026). No method published. gartner.com
- 11 Gartner, press release, "Over 40% of agentic AI projects will be canceled by end of 2027" (25 Jun 2025). gartner.com
- 12 KPMG US, Quarterly AI Pulse, Q3 2026 (24 Sep 2026). n = 314, US firms of \$1bn+ revenue. kpmg.com
- 13 Bain, "Executive Survey: AI Moves from Pilots to Production" (24 Nov 2025). n = 197. bain.com
- 14 Menlo Ventures, *2025: The State of Generative AI in the Enterprise* (9 Dec 2025). n = 495, US. Menlo invests in Anthropic. menlovc.com
- 15 Challapally et al., *The GenAI Divide: State of AI in Business 2025*, MIT NANDA (Jul 2025, released 18 Aug 2025). 52 interviews, 153 survey responses. Now by request at media.mit.edu
- 16 BCG, *The Formula for Agentic AI Value*, Applied AI Index 2026 (30 Sep 2026). n = 1,330. bcg.com
- 17 Deloitte, "AI ROI: The paradox of rising investment and elusive returns" (22 Oct 2025). n = 1,854, Europe and Middle East. deloitte.com
- 18 McKinsey, *The state of AI: How organizations are rewiring to capture value* (12 Mar 2025). n = 1,491, fielded Jul 2024. mckinsey.com
- 19 LangChain, *State of Agent Engineering* (Dec 2025). n = 1,340, self-selected. LangChain sells evaluation tools. langchain.com
- 20 Gartner, press release, "Regular AI system assessments triple the likelihood of high GenAI value" (4 Nov 2025). n = 360. gartner.com
- 21 KPMG US, Quarterly AI Pulse, Q2 2026 (24 Jun 2026). n = 204, US firms of \$1bn+ revenue. kpmg.com
- 22 Bain, "Your AI Budget Is Growing. Your Returns Aren't." (1 Jun 2026). Automation and AI Pathfinder Survey, n = 951. bain.com
- 23 Grant Thornton, *2026 AI Impact Survey* and mid-market analysis (2026). n = 950, Feb–Mar 2026. grantthornton.com
- 24 Gartner, press release, "AI inference costs per agentic workflow will increase more than fivefold through 2028" (17 Aug 2026). gartner.com
- 25 RSM US, *Middle Market AI Survey 2026* (21 Jul 2026). n = 1,030, US and Canada, firms of \$30m–\$10bn revenue, all AI users. rsmus.com
- 26 Gartner, press release on AI agent governance (26 May 2026). gartner.com
- 27 Gartner, "Lack of AI-Ready Data Puts AI Projects at Risk" (26 Feb 2025). gartner.com
- 28 KPMG US, Quarterly AI Pulse, Q1 2026 (31 Mar 2026). n = 237, US firms of \$1bn+ revenue. kpmg.com
- 29 IBM Institute for Business Value, 2026 CEO Study, press release (4 May 2026). n = 2,000 CEOs. ibm.com
- 30 Gartner, press release, "Successful AI initiatives invest up to four times more in data and analytics foundations" (16 Apr 2026). n = 353. gartner.com
- 31 Gartner, press release on AI maturity and three-year production lifetimes (30 Jun 2025). n = 432. gartner.com
- 32 RSM US, 2025 Middle Market AI Survey, press release (11 Jun 2025). n = 966, US and Canada. rsmus.com
- 33 Gartner, press release, "Worldwide AI spending to grow 49.5% in 2026" (16 Sep 2026). gartner.com
- 34 Gartner, press release on LLM inference costs to 2030 (25 Mar 2026). gartner.com
- 35 DoIT, "Why 79% of Enterprises Overspent on AI in 2026" (9 Jun 2026). Sapio Research, n = 500, US and UK, Feb 2026. Vendor-commissioned. doit.com
- 36 Gartner, "Gartner Identifies Three Pillars for Deriving Value from AI" (9 Mar 2026). n = 353. gartner.com
- 37 Andreessen Horowitz, "Leaders, gainers and unexpected winners in the Enterprise AI arms race" (30 Jan 2026). n = 100, Global 2000. a16z invests in OpenAI. a16z.com
- 38 Model deprecation pages, accessed 2 Oct 2026: [Anthropic](https://anthropic.com), [OpenAI](https://openai.com), [Google Cloud](https://google.com).
- 39 Hugging Face, "Security incident disclosure, July 2026" (16 Jul 2026) and technical timeline (27 Jul 2026); OpenAI statement (21 Jul 2026). huggingface.co
- 40 Microsoft MSRC, CVE-2025-32711 (Jun 2025); Reddy and Gujral, "EchoLeak", arXiv:2509.10540. microsoft.com
- 41 OWASP, *Top 10 for LLM Applications 2025* (Nov 2024). genai.owasp.org
- 42 NCSC, "Prompt injection is not SQL injection (it may be worse)" (8 Dec 2025). ncsc.gov.uk
- 43 IBM, *Cost of a Data Breach Report 2026*, press release and X-Force analysis (29 Jul 2026). 602 breached organisations. ibm.com
- 44 UK Parliament, written statement HCWS314, "Artificial intelligence update" (7 Sep 2026). parliament.uk
- 45 EU: Regulation (EU) 2024/1689 (AI Act), incl. Articles 2, 15, 99, 113; Regulation (EU) 2026/1744, the Digital Omnibus on AI (OJ 24 Jul 2026). eur-lex.europa.eu
- 46 UK: King's Speech 2026 briefing notes (13 May 2026); Data (Use and Access) Act 2025, s.80 and S.I. 2026/82; ICO guidance status page; DSIT response on AI regulation (6 Feb 2024). legislation.gov.uk
- 47 US: Executive Order 14365 (11 Dec 2025); White House legislative recommendations (20 Mar 2026); Colorado SB26-189 (signed 14 May 2026). leg.colorado.gov

NEXT STEP

Start with a briefing, not a sales call.

If you have a pilot that half-works and won't ship, or a board asking what the AI budget is buying, we'll spend 30 minutes on your situation and give you a straight answer: what it would take to get into production, and whether it's worth doing.

[REQUEST AN EXECUTIVE BRIEFING → KISEKILABS.COM/BOOK-A-CALL](https://kisekilabs.com/book-a-call)

Not ready for a conversation? Take the AI Readiness Assessment at kisekilabs.com/ai-readiness-assessment to see where your organisation stands before your next steering committee.

WHERE WE FIT

Opportunity Diagnostic	Pilot to Production	Advisory Retainer	Team Enablement
2 TO 3 WEEKS	4 TO 8 WEEKS	MONTHLY	4-WEEK COHORT
The mandate is clear but the target isn't. We find the first workflow worth the money and size the bet.	Something half-works and needs to ship. Evaluation suite first, then the build, then a complete handover.	AI decisions are frequent and expensive. A senior challenger on vendors, architecture and budget.	Your teams need to use AI, not be briefed about it.

ABOUT KISEKI LABS

AI advisory and delivery for mid-market and enterprise organisations in the UK and US: AI that survives contact with production. Founder Eddie Forson has built production AI systems since 2017, from computer vision to LLM agents, much of it inside regulated financial services. Every engagement has a fixed scope, a fixed fee and a defined end date. kisekilabs.com

Kiseki Labs engagements: "[From 60% to 98%](#)" and the KYC case on kisekilabs.com. Eddie Forson's quotations are from his LinkedIn newsletter *AI for Decision Makers* (June and September 2026).