David Soden

MARKET ASSESSMENT | AUGUST 2026

The Automation Platform Overhaul

What the license actually buys, what AI-assisted build changes, and the measurement that makes either one defensible

David Soden

Independent analyst and practitioner, enterprise automation and applied AI

davidsoden.com/market-assessment

REPORT DS-2026-CDE7 · 2026 · VERSION 1.0 · PUBLIC

About this report

Every substantive claim in this report carries an evidence classification and, where applicable, a numbered source. Facts are separated from vendor claims, third-party estimates, the author's assessments, and untested hypotheses. Forward-looking sections use scenarios with observable tripwires rather than point forecasts. Stated assumptions are listed before the analysis begins.

Market Assessment reports are published and commissioned at **davidsoden.com/market-assessment**. This document is analysis, not investment, legal, or procurement advice.

Scope and evidentiary standard

This assessment covers one decision: whether an organization already running automation on a commercial platform should rebuild that capability on infrastructure it owns, with AI-assisted development doing the construction. It is written to be industry agnostic. The economics, the control requirements, and the measurement problem are the same in telecommunications, insurance, logistics, and public sector work, because the thing being priced is engineering labor and evidence production, not domain knowledge.

Four questions organize the analysis. What does the platform license actually pay for, once you separate the runtime from everything wrapped around it. What does AI-assisted construction change about the cost of building automations, and what does it leave untouched. What has to exist in a governance and control layer before an internal auditor, an external auditor, or a regulator will accept the output. And how do you measure the result against a baseline that existed before the automation did, given that almost nobody captures one.

Published numbers in this category need discounting for three specific reasons, and the discount is large.

First, list prices stop at the entry tier. Microsoft publishes per-user and per-bot pricing for Power Automate. UiPath publishes a single figure, twenty-five dollars a month for its Basic tier, and routes everything above it to a sales conversation. Any cost comparison at enterprise scale is therefore a comparison against a number the vendor declines to publish, and the discount off list at the top of the market is not observable from outside.

Second, most of the quantitative research on AI-assisted development is produced by companies selling something adjacent to the finding. GitClear sells code-quality analytics and publishes research showing code quality falling. Veracode sells application security testing and publishes research showing AI-generated code is insecure. DORA is funded by Google Cloud. None of that

makes the findings wrong. It does mean the effect sizes are the least reliable part and the direction is the most reliable part, and this report treats them that way.

ASSESSMENT Third, the single most repeated statistic in enterprise AI, the claim that ninety-five percent of generative AI pilots produce no return, comes from a preliminary, non-peer-reviewed MIT Media Lab working paper built on 150 interviews, 350 survey responses, and 300 observed deployments. It is cited across the trade press as though it were a census. It is not, and no conclusion in this report rests on it. [18]

Every claim below carries one of six labels. Where a claim could not be sourced to something a reader can open and check, it is labeled as the author's assessment or dropped.

Label	Definition	How to treat it
FACT	Verifiable in the cited primary source: a filing, press release, official documentation, or standards body announcement.	Reliable as stated. Check the source if the exact wording matters.
VENDOR CLAIM	Reported by a vendor about its own business or product. Self-measured and not audited by any third party.	The vendor said it. Directionally useful, not a measurement. Definitions of the metric are usually undisclosed.
THIRD-PARTY ESTIMATE	A research firm forecast or survey result. Methodology is proprietary and generally not reproducible.	Use for direction and order of magnitude only. Do not build a model on it.
ASSESSMENT	The author's reasoned judgment from the cited evidence. Falsifiable, and the reasoning is shown so it can be argued with.	This is analysis, not reporting. Disagree with it where your evidence differs.
HYPOTHESIS	A proposition offered for testing. Not currently supported by sufficient evidence to assert.	Treat as a research question, not a conclusion. Each one names what would confirm or refute it.
SCENARIO	A structured future state with a stated subjective probability. The probability is the author's judgment, not a calculated figure.	Use the leading indicators attached to each, not the probability. The indicators are observable; the probability is not.

Assumptions this analysis rests on

These are premises, not findings. Each one is stated with the part of the analysis that fails if it turns out to be wrong.

The organization can hire and retain three to six engineers who are competent to run production infrastructure, not just to write application code. If this is wrong, the entire build path collapses and the recommendation reverses to buying, because an owned control plane with nobody qualified to operate it is worse than a rented one.

Model capability for code generation holds at roughly its current level and inference prices do not rise sharply. If this is wrong, the build column of the cost model is understated, and the gap widens in favor of the licensed platform.

Regulators and auditors continue to treat an automated decision as the responsibility of the organization deploying it, regardless of which vendor's software executed it. If this is wrong, and liability shifts meaningfully onto platform vendors, the compliance argument for owning your own evidence trail weakens considerably.

Published list prices approximate what mid-market buyers pay. Large enterprises negotiate, sometimes steeply. If this is wrong at your scale, the break-even calculation in the cost section moves against building, because the saved license is smaller than modeled.

The processes being automated are stable enough that a measurement taken today still describes them in twelve months. If this is wrong, baselining measures noise and the ROI ledger becomes theater.

Open-core tools chosen for the owned stack keep their current licenses. If this is wrong, and a relicensing event occurs, the owned stack acquires a migration cost that was never priced. This has happened repeatedly in adjacent infrastructure categories and should be treated as a live risk rather than a tail one.

The call

ASSESSMENT Stop framing this as build versus buy and decide it layer by layer. The rent worth cancelling is the per-bot execution license, which prices a commodity. The rent worth paying is the connector library and the vendor's compliance attestations, at least until your own control plane can produce the same evidence. What makes an owned stack defensible is not that AI wrote it faster; it is the measured baseline, the per-run cost ledger, and an audit trail an outsider can reproduce without your help. Build those three first and the platform choice stops mattering. I would reverse this if AI-assisted code proved cheaper to maintain than to write, and the evidence currently points the other way.

My judgment on the evidence assembled here, nothing more. There is data I can't see and data I may have missed. New evidence can move me, including to the opposite position, and I reserve the right to move.

Executive summary

The argument for cancelling platform rent is usually made on the license line and usually loses on the labor line. The argument for keeping the platform is usually made on governance and usually understates how much governance you still have to build yourself. Both sides are arguing about the wrong layer.

FINDING 1 The license is rarely the dominant cost, and the people making the build case rarely model the alternative honestly

THIRD-PARTY ESTIMATE A mid-size program of forty unattended automations and 250 licensed makers costs roughly $117,000 a year in Power Automate list-price licensing. Two additional engineers cost about three times that. Under the cost model in this report, the owned stack runs roughly a quarter more expensive over five years than the licensed one, because engineering labor dominates both columns and the owned path needs more of it. The build case wins on scale or on strategic control, not on the license line. [5]

FINDING 2 Cancelling the platform does not cancel the control plane, it transfers it to you

ASSESSMENT What a commercial automation platform sells is not mainly execution. It is a credential vault, a connector library, a scheduler with retry semantics, role-based access control, an immutable run log, tenant isolation, a support contract, and a stack of third-party attestations. Rebuilding execution is a weekend. Rebuilding the rest, to a standard an auditor accepts, is the actual project, and it is the part that AI-assisted coding helps least with. [27] [28]

FINDING 3 AI-assisted construction lowers the cost of the first version and raises the cost of the tenth

ASSESSMENT Google's DORA program found near-universal AI adoption among developers alongside a negative relationship between AI use and delivery stability, describing AI as an amplifier of whatever system it lands in. GitClear's telemetry over 623 million code changes shows duplicated blocks up 81 percent and refactoring line moves down 70 percent since 2023. Veracode found models chose the insecure option in 45 percent of security-relevant coding tasks, with no improvement as models got larger. Speed of first draft is real. Cost of ownership is where it comes back. [14] [16] [17]

FINDING 4 Break-even against license spend sits at roughly 140 to 200 unattended automations

THIRD-PARTY ESTIMATE At Microsoft's published $150 per bot per month for unattended RPA, 200 automations pay for two additional engineers. At the $215 hosted rate, 140 do. Below that count the owned stack is a strategic choice you are funding, not a saving you are booking, and it should be defended as such in the business case rather than dressed up as cost avoidance. [5]

FINDING 5 The incumbent's growth has already flattened, which is a negotiating fact before it is a strategy fact

FACT UiPath closed fiscal 2026 on 31 January 2026 with $1.853 billion in ARR, up 11 percent year over year, full-year revenue of $1.611 billion up 13 percent, and a dollar-based net retention rate of 107 percent. A retention rate that close to flat means the installed base is expanding barely faster than it is contracting. Vendors in that position discount to protect renewals. [1] [2] [3] [4]

FINDING 6 The observability standard you would build an owned control plane against is not finished

FACT As of the v1.42.0 release on 12 June 2026, OpenTelemetry moved every gen_ai.* attribute, span, and metric out of the main semantic-conventions repository into a dedicated

GenAI conventions repository. That is a release-cadence change, not a graduation. Nothing in the GenAI conventions is marked stable. Anyone instrumenting an agent platform today is building against a moving target and should budget for re-instrumentation. [25] [26]

FINDING 7 The open-source governance layer consolidated under a database vendor in January 2026

FACT ClickHouse acquired Langfuse on 16 January 2026 alongside a $400 million Series D that valued ClickHouse at $15 billion. ClickHouse committed publicly to keeping the MIT license on core features and keeping self-hosting a first-class path. The commitment is real and it is also unenforceable, which is the standard condition of open-core dependency and should be priced as such. [9] [10] [11]

FINDING 8 Almost nobody captures the baseline the entire business case depends on

THIRD-PARTY ESTIMATE IBM's CEO study found roughly 25 percent of AI initiatives delivered the expected return and 16 percent had scaled enterprise-wide. IBM's later work puts the share of executives who can confidently measure AI return at 29 percent. These are self-reported executive surveys with unpublished instruments, so the precision is weak and the direction is not. Without a pre-automation baseline there is no attribution, only testimonial. [19] [20]

FINDING 9 Agent and bot identity is the control gap an auditor will find first

THIRD-PARTY ESTIMATE Vendor-sponsored counts put non-human identities at 45 to 1 against humans in the average enterprise, 80 to 1 by KPMG's figure, and 144 to 1 in cloud-native environments. The definitions differ and none of the instruments are published. What is consistent is the governance gap: roughly nine in ten organizations say their identity tooling cannot manage AI agent identities, and fewer than half have any policy for them. [29] [30]

FINDING 10 The regulatory clock moved in two directions at once during 2026

FACT The EU AI Act's Article 50 transparency obligations became applicable and enforceable on 2 August 2026, covering systems that interact with people or generate synthetic content regardless of risk classification. In the same period the AI Omnibus deferred the high-risk obligations to 2 December 2027. Deferral is not repeal, and building the evidence trail on the original timetable remains the cheaper option. [21] [22] [23]

The category is four markets wearing one name

The premise that UiPath, Power Automate, and n8n are alternatives to each other, and jointly an alternative to writing code, is where most of these programs go wrong. They occupy different layers, and only one of those layers is genuinely commoditized.

ASSESSMENT Deterministic task automation is the RPA layer: screen and API interaction with legacy systems that have no usable interface. Workflow and integration is the connector layer: moving structured data between SaaS products on a trigger. Durable orchestration is the execution layer: guaranteeing that a long-running, multi-step process survives a process restart, a network partition, and a deploy. Agent runtime and governance is the newest layer: model calls,

tool selection, prompt versioning, evaluation, and trace capture. A single vendor sells across several of these, which is why buyers experience them as one market.

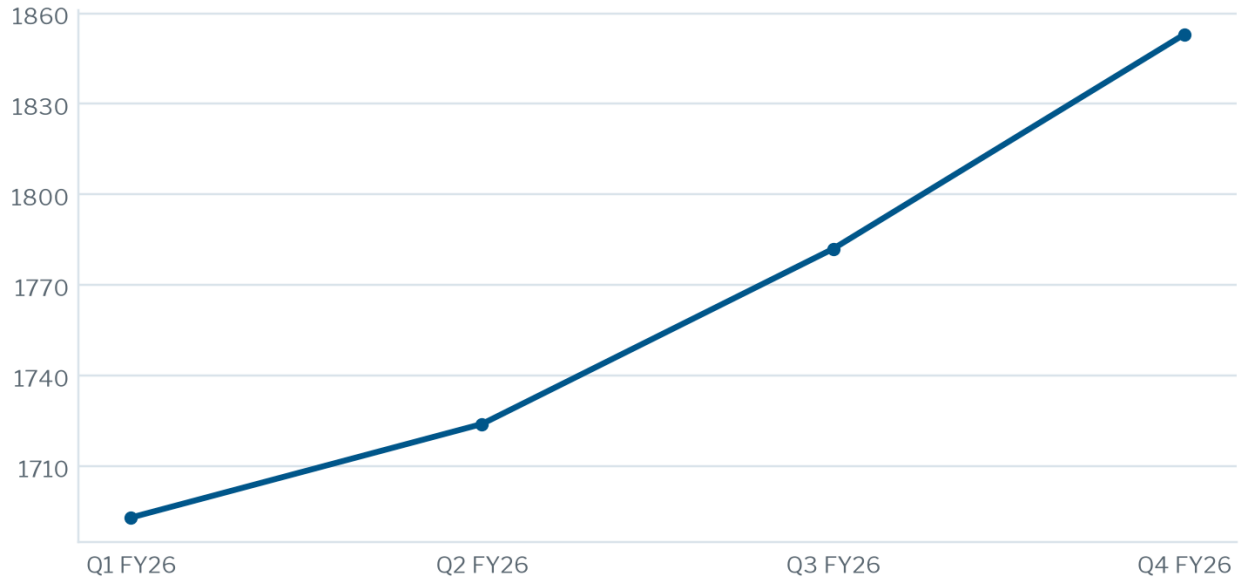
FACT Forrester's automation analysts describe the individual markets, RPA, integration platform as a service, and business process management, as having all but converged, with the enterprise automation space moving toward AI-led orchestration. Their 2026 prediction set, published 3 November 2025, forecasts that fewer than 15 percent of firms will turn on the agentic features in their intelligent automation suites during 2026. [13]

ASSESSMENT That 15 percent figure is the most useful number in the vendor conversation, and it cuts against the vendors who published the features. Buyers are paying for agentic capability at renewal and not switching it on. That is a pricing lever, not a technology finding.

Layer	What it actually does	Rent or own	What you take on if you own it
Deterministic task automation	Drives UIs and APIs of systems with no integration surface, including terminal and desktop applications	Rent, usually	Screen-scraping maintenance, OS image management, per-application breakage on every vendor update
Workflow and integration	Moves structured data between SaaS products on triggers and schedules	Own, increasingly	Authoring and maintaining connectors, credential rotation, rate-limit and pagination handling per API
Durable orchestration	Guarantees long-running processes survive restarts, partitions, and deploys	Either, and the open options are strong	State store operation, versioning of in-flight workflows, replay determinism
Agent runtime and governance	Model calls, tool selection, prompt versioning, evaluation, trace capture	Own the control plane	Trace schema drift, evaluation harness, prompt release process, model deprecation churn

ASSESSMENT The layer worth owning is not the one that is hardest to build. It is the one where your requirements diverge most from every other buyer's, which is governance, and the one where switching cost is highest if you get it wrong, which is also governance. Execution is where teams start because it is visible and satisfying, and it is the least valuable thing to own.

FIGURE 1 UiPath annual recurring revenue by quarter, fiscal 2026 ($ millions)



FACT Q1, Q3, and Q4 figures are reported directly in the company's quarterly releases. The Q2 figure is derived by adding reported net new ARR of $31 million to the Q1 balance, because the company reported net new ARR rather than the balance that quarter. Growth across the year ran 11 to 12 percent with dollar-based net retention between 107 and 108 percent. [1] [2] [3] [4]

What you are actually renting

The instinct that per-bot licensing prices a commodity is correct. The conclusion that the whole license is therefore waste does not follow, because the runtime is not the expensive part of what the vendor ships.

FACT Microsoft's published list prices are $15 per user per month for Power Automate Premium, which covers cloud flows and attended desktop flows, $150 per bot per month for Power Automate Process, which covers unattended RPA, and $215 per bot per month for Hosted Process, which adds a Microsoft-managed virtual machine. The unit of charge changes from the user to the concurrent bot at the moment a process runs unattended. [5]

FACT UiPath publishes a single figure on its pricing page, Automation Cloud Basic starting at $25 per month, and routes Standard, Enterprise, and both Test Cloud tiers to a sales conversation with no published rate. [6]

FACT n8n publishes tenant-level pricing billed by workflow execution rather than by step: Starter at 20 euros per month for 2,500 executions, Pro at 50 euros for 10,000, Business at 667 euros for 40,000, and Enterprise on request. The self-hosted community edition is free and enforces no execution limit; self-hosted Business and Enterprise require a license key and phone home to a license server daily. [7]

FACT Temporal Cloud starts at $100 a month for one million Actions with 1 GB of active storage, with active storage billed at $0.042 per gigabyte-hour and retained storage at $0.00105. The self-hosted server is open source and free. [8]

FIGURE 2 **Published monthly list price, and what the unit of charge actually is**

FACT US dollars per month, billed annually, as published on each vendor's own pricing page in August 2026. The units are deliberately not normalized, because the change of unit is the finding: a per-user price and a per-concurrent-bot price are not comparable, and the switch between them is what makes unattended automation expensive. n8n is excluded because it publishes in euros; its figures appear in the text. Every tier above the entry point at UiPath is quote-only. [5] [6] [8]

ASSESSMENT Strip the runtime out of those prices and what remains is a list of things that are boring to build and painful to operate: a credential vault with rotation, several hundred maintained connectors, a scheduler with retry and backoff semantics that survive partial failure, tenant isolation, role-based access control, an immutable execution log, a support contract with a response time in it, and SOC 2, ISO 27001, and increasingly ISO 42001 attestations that your auditor already knows how to read.

ASSESSMENT The connector library is the single most underestimated line. A connector is not an HTTP client. It is authentication flows, token refresh, pagination, rate-limit backoff, schema drift handling, and a maintenance commitment for every upstream API version bump, multiplied by every system in the estate. AI-assisted development writes the first version of a connector quickly and does nothing about the maintenance commitment, which recurs forever and scales with the number of integrations rather than the number of automations.

FACT Microsoft's own governance tooling illustrates how much of this is operational rather than technical. The Power Platform Center of Excellence Starter Kit, the reference governance implementation Microsoft shipped for years, is no longer actively maintained; Microsoft's documentation now states that issues are no longer reviewed or addressed and directs administrators to native Power Platform admin center capabilities instead. [27]

ASSESSMENT That deprecation is worth reading carefully by anyone planning to build a governance dashboard. Microsoft built the reference version, maintained it for years with a dedicated team, and still retired it into the product because a bolt-on governance layer built on

the same low-code substrate it governs is expensive to keep current. An internally built hub over a heterogeneous automation estate faces the same gravity with less funding.

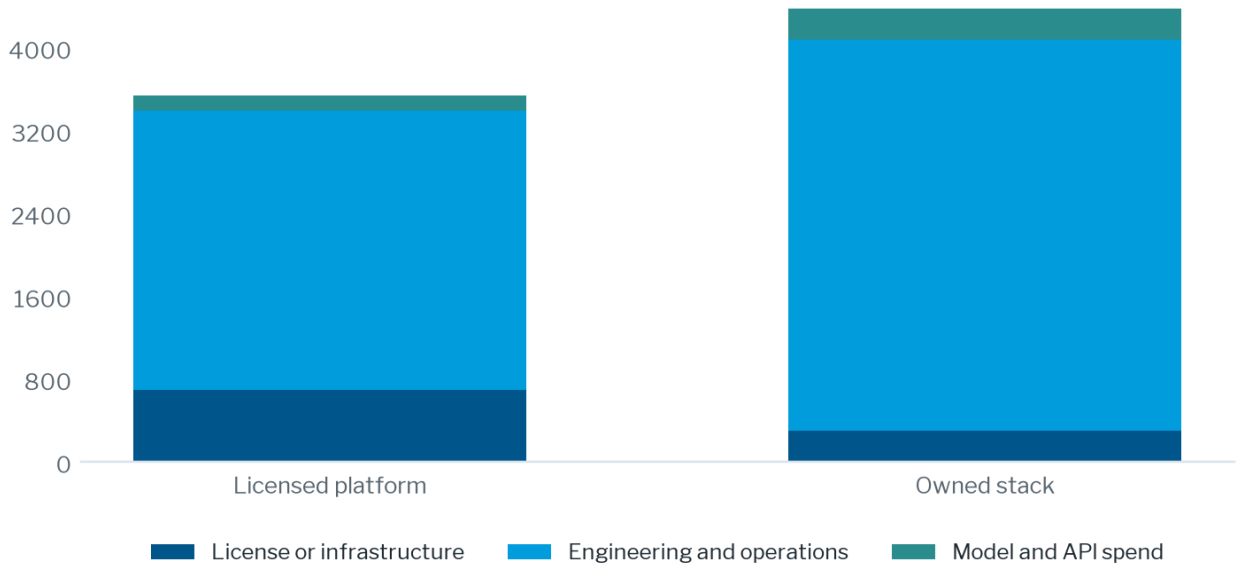
The five-year cost, modeled honestly

The following model is an estimate with stated assumptions, not a finding. It is included because most build cases in this category compare a license line to zero and then discover the difference in headcount two years later.

THIRD-PARTY ESTIMATE Assume forty unattended automations and 250 licensed makers, which is a realistic mid-size program. At Microsoft list prices that is $72,000 a year in Process licensing plus $45,000 in Premium seats, or $117,000 a year, growing 8 percent annually. Assume the licensed path also needs three loaded engineers and administrators at $180,000 each, because a platform still needs people to run it, and $30,000 a year of model and API spend. Five-year total: approximately $3.54 million. [5]

THIRD-PARTY ESTIMATE For the owned stack, assume five engineers in year one to build and four thereafter, at the same loaded rate, $60,000 a year of infrastructure, and $60,000 a year of model and API spend, reflecting both development and runtime inference. Five-year total: approximately $4.38 million, or about 24 percent more than the licensed path.

FIGURE 3 **Five-year total cost by path, mid-size program ($ thousands)**



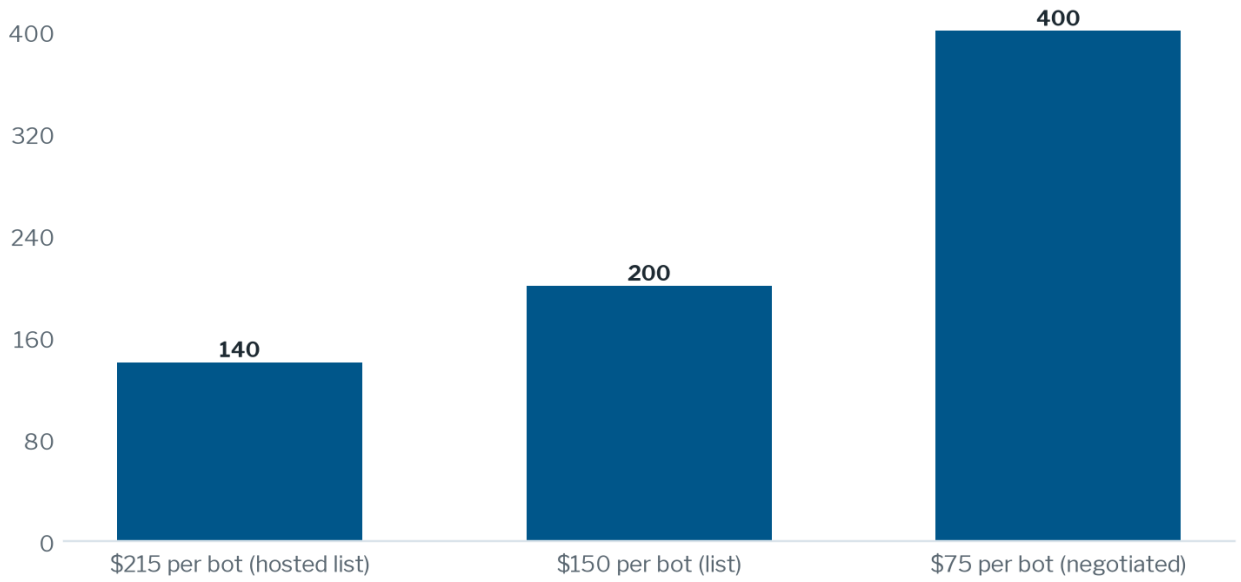
THIRD-PARTY ESTIMATE The author's model, built on the published list prices in this report and stated headcount assumptions of three loaded engineers on the licensed path and five falling to four on the owned path, at $180,000 loaded each. Loaded cost per engineer is the most sensitive input; at $250,000 the gap widens, at $120,000 it narrows to near parity. This model assumes no negotiated discount, which favors the owned path, and no platform migration cost, which also favors it. [5]

ASSESSMENT The direction of that result is more durable than the magnitude. In a mid-size program, license spend is a rounding error against engineering payroll, and the owned path needs more payroll rather than less, because you have taken on the vendor's maintenance obligations

along with their runtime. Anyone whose build case shows a large saving should check whether the model includes the second and third year of maintaining what year one produced.

THIRD-PARTY ESTIMATE The break-even is a bot count, and it is a useful number to carry into a renewal negotiation. Two additional engineers cost roughly $360,000 a year. At $150 per bot per month, that is 200 unattended automations. At the $215 hosted rate, 140. At a negotiated $75, 400. [5]

FIGURE 4 Unattended automations required for license spend to equal two additional engineers



THIRD-PARTY ESTIMATE The author's calculation, dividing $360,000 of annual loaded cost for two engineers by the annualized per-bot price. It assumes the owned stack needs exactly two more engineers than the licensed one and that automation count maps one to one onto bot licenses, which overstates the case where several automations share a bot. Below these counts an owned stack is a strategic choice, not a saving. [5]

ASSESSMENT This is where the original build thesis needs amending rather than defending. The strongest version of the argument is not that platform rent is ridiculous. It is that platform rent is small, which means it should never have been the reason for the decision in the first place, and that the money at stake is in the engineering organization either way. Deciding on the license line is how programs end up owning a control plane nobody budgeted to operate.

What AI-assisted construction actually changes

The claim that AI-assisted coding makes building this stack fast is true and incomplete. The evidence base on what happens after the first draft is now large enough to argue from, and it is not flattering.

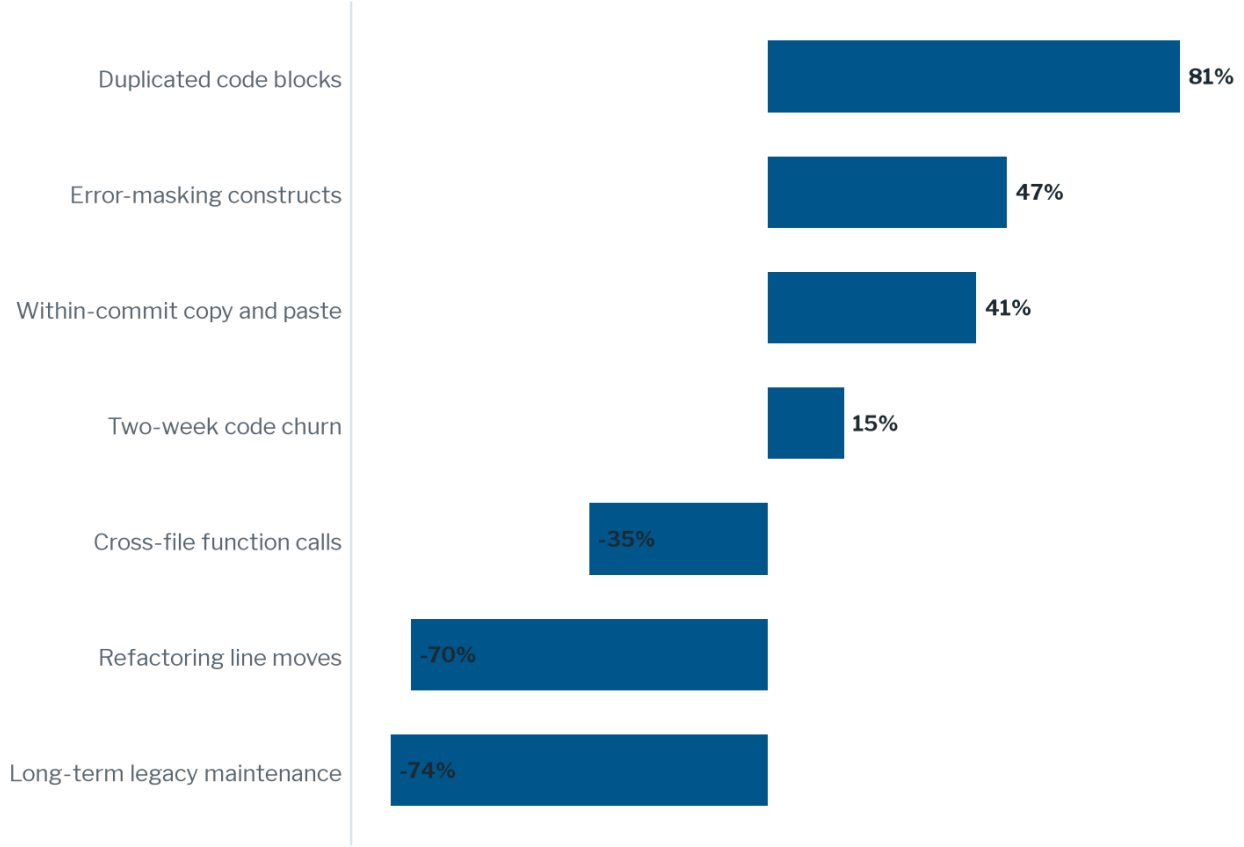
FACT Google's DORA research program reported near-universal adoption, with 90 percent of respondents using AI in software development and over 80 percent believing it increased their productivity, alongside continued low trust: roughly 30 percent did not trust AI-generated code and under a quarter trusted it a lot. The same research found AI adoption maintains a negative relationship with software delivery stability, and characterized AI's role as an amplifier that

magnifies both the strengths of high-performing organizations and the dysfunctions of struggling ones. [14]

FACT DORA's follow-on work in 2026 on the return from AI-assisted development places the constraint downstream of code generation, arguing that faster authoring produces value only when review, testing, and deployment can absorb the increased change volume, and describing a J-curve in which value arrives after a dip rather than immediately. [15]

VENDOR CLAIM GitClear, which sells code-quality analytics and therefore has a commercial interest in the finding, analyzed 623 million code changes from 2023 to 2026 and reported duplicated code blocks up 81 percent, within-commit copy and paste up 41 percent, error-masking constructs up 47 percent, and two-week churn up 15 percent, while cross-file function calls fell 35 percent, refactoring line moves fell 70 percent, and long-term legacy maintenance fell 74 percent. [16]

FIGURE 5 **Change in code maintainability signals, 2023 to 2026 (percent)**



VENDOR CLAIM GitClear's own telemetry across 623 million analyzed code changes, published January 2026. GitClear sells code-quality analytics, so treat the effect sizes as the vendor's and the direction as the signal. The sample is drawn from repositories using GitClear, which is self-selected and skews toward organizations already worried about code quality. The measurement does not isolate AI as the cause; it establishes correlation across the period of AI adoption. [16]

VENDOR CLAIM Veracode, which sells application security testing, evaluated more than 100 models across 80 curated coding tasks and found that when a task offered both a secure and an insecure implementation, models chose the insecure one 45 percent of the time. Java performed

worst at over 70 percent failure; Python, C#, and JavaScript fell between 38 and 45 percent. The finding the buyer should carry is the trend line: functional correctness improved with model scale and security performance stayed flat. [17]

ASSESSMENT Read together, these three sources describe an economics that is the opposite of the usual pitch. AI-assisted development compresses time to first working version dramatically and leaves the cost of ownership at least where it was, plausibly higher. For a program building forty automations that will each live five years and be modified quarterly, first-version cost is a small share of lifetime cost. The productivity claim is real and it is being applied to the wrong denominator.

ASSESSMENT The scale of spending on AI-assisted coding is worth citing for what it does not establish. Press reporting sourced to unnamed people familiar with the matter puts Anthropic's Claude Code at a run rate in the billions and at more than half of that company's enterprise revenue, none of it disclosed by the company. Rapid adoption of a development tool is evidence that developers want it. It says nothing about maintenance cost, which is the number this decision turns on, and adoption figures are routinely offered as though it did. [35]

ASSESSMENT The practical consequence is that AI-assisted build is a strong argument for owning the control plane, which is written once and maintained deliberately by a small team, and a weak argument for owning hundreds of individual automations, which are written constantly by many hands and where duplication and error-masking compound. That inverts how most teams sequence the work.

The control plane is the product

The hub-and-spoke instinct, one application above the automations that carries logging, access control, and cost, is the right shape. The version below is that shape with the parts that get skipped added back, because the skipped parts are the ones an auditor asks about.

ASSESSMENT A control plane worth building has seven components, and only two of them are dashboards. A registry of automations, each with an owner, a criticality rating, and a linked process baseline. A prompt and configuration registry with versioning and a promotion path, so that changing a prompt is a release rather than an edit. A model gateway that terminates every model call, so that spend, latency, and provider choice are policy rather than per-automation code. A trace store capturing execution, tool calls, inputs, outputs, and cost per run. An identity layer that issues a distinct credential per automation, with an owner and an expiry. A policy engine enforcing which automations may touch which data classes. And an evidence exporter that produces auditor-ready packets without an engineer writing a query.

ASSESSMENT Langfuse is a reasonable choice for the trace store and the prompt registry, and it is worth being specific about why: it is MIT-licensed at the core, framework-agnostic, and it runs self-hosted at production scale rather than as a limited demo. It is not, on its own, a control plane. It observes; it does not authorize, meter, or enforce. Treating an observability tool as the governance layer is the most common architectural error in this pattern, and it is discovered during the first audit rather than during design. [11]

FACT Langfuse's independence changed in January 2026. ClickHouse acquired the company on 16 January 2026, alongside a $400 million Series D that tripled ClickHouse's valuation to $15 billion, and committed publicly to keeping the MIT license on core features, keeping self-hosting a first-class option, and leaving the roadmap unchanged. Langfuse reported more than 20,000 GitHub stars, 23.1 million monthly SDK installs, 6 million Docker pulls, and use by 19 of the Fortune 50. [9] [10]

ASSESSMENT The commitment reads as sincere and it is not enforceable. An MIT license cannot be revoked on code already published, so the floor is solid; the roadmap, the enterprise features, and the release cadence are not covered by that floor. The correct response is not to avoid the tool. It is to keep the trace schema and the exported evidence in a format you control, so that a future relicensing or roadmap change costs you an adapter rather than a migration.

FACT The vendor-neutral standard that would make that portability real is not finished. OpenTelemetry moved all gen_ai.* attributes, spans, and metrics out of the main semantic-conventions repository into a dedicated GenAI conventions repository at the v1.42.0 release on 12 June 2026, giving that work its own release cadence. The conventions now model an agent run as a span tree covering create_agent, invoke_agent, invoke_workflow, execute_tool, retrieval, and plan operations, and MCP tool calls share the same vocabulary. Nothing in the set is marked stable. [25] [26]

ASSESSMENT Instrument to the GenAI conventions anyway, and budget for one re-instrumentation pass before they stabilize. The alternative, a proprietary trace schema, guarantees the migration cost instead of risking it. The moving target is cheaper than the dead end.

Component	Open-source option	Commercial option	Why it is not optional
Trace and evaluation store	Langfuse (MIT core)	Langfuse Cloud, LangSmith, Braintrust	Without per-run traces there is no evidence and no unit cost
Prompt and config registry	Langfuse prompt management	Portkey	An unversioned prompt is an unversioned business rule
Model gateway	LiteLLM (MIT)	Portkey, Kong AI Gateway	Terminates spend, routing, and provider risk in one place
Durable execution	Temporal, Restate, Inngest	Temporal Cloud, AWS Step Functions, Cloudflare Workflows	Retry and replay semantics you would otherwise reimplement badly
Automation registry and RBAC	Build on your existing IdP	Platform-native (Power Platform admin center)	Ownership and access review are the first two audit questions
Non-human identity	IdP service principals with rotation	CyberArk, Entro, Astrix	One shared service account across automations destroys attribution

Component	Open-source option	Commercial option	Why it is not optional
Process baseline capture	Instrumented ticket and event logs	Celonis, UiPath Process Mining, SAP Signavio	No baseline means no measured return, only a claimed one

FACT Temporal raised a $300 million Series D at a $5 billion valuation on 17 February 2026, led by Andreessen Horowitz, and repositioned explicitly around durable execution for agentic applications, reporting 380 percent revenue growth and 350 percent growth in weekly active usage over the prior year. Growth rates stated as percentages from an undisclosed base establish direction and nothing about scale. [33]

FACT n8n raised $180 million in October 2025 at a $2.5 billion valuation led by Accel, with NVIDIA's venture arm participating. Its self-hosted community edition remains free with no execution limit, while self-hosted Business and Enterprise editions require a license key that contacts a license server daily. [34] [7]

ASSESSMENT That daily license check is the detail worth noting in an air-gapped or high-assurance environment, and it is the kind of thing that does not appear in a build-versus-buy spreadsheet until deployment. Self-hosted is not the same as self-contained.

FACT The Model Context Protocol, which is how most agent tooling now exposes capabilities, has a published security profile from a national authority: the NSA issued a Cybersecurity Information Sheet covering MCP risks. Independent analysis in 2026 puts the public registry past 5,000 servers, with tool-description poisoning, prompt injection through tool metadata, and context leakage as the recurring failure classes. [24]

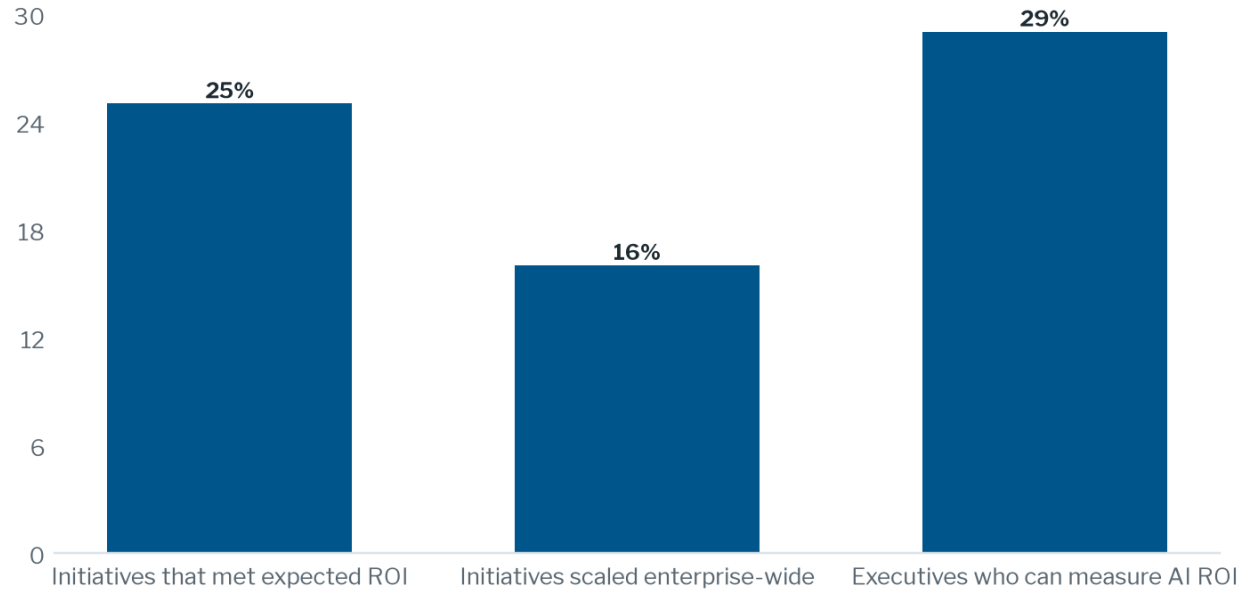
ASSESSMENT For a regulated estate, the workable rule is a curated internal MCP registry with security review as a gate, installation from arbitrary URLs blocked at the endpoint, and every tool call traced with the agent identity that issued it. That is three controls, all cheap, and all much harder to retrofit than to install.

Baseline before automation, and why it is the whole argument

The strongest idea in the original thesis is also the least common in practice: measure the process before you automate it, and keep measuring after, so that the saving is calculated rather than asserted. It deserves a stronger treatment than it usually gets, because the naive version produces numbers that fall apart under audit.

THIRD-PARTY ESTIMATE IBM's CEO study across 2,000 chief executives found roughly 25 percent of AI initiatives had delivered the expected return and 16 percent had scaled enterprise-wide. IBM's subsequent work puts the share of executives who can confidently measure AI return at 29 percent, with the missing or imprecise pre-deployment baseline named as the usual cause. Both are self-reported executive surveys with unpublished instruments, so the numbers are directional. [19] [20]

FIGURE 6 **Self-reported enterprise AI return, four measures (percent)**



THIRD-PARTY ESTIMATE Drawn from IBM Institute for Business Value work. These are three different questions asked of three overlapping executive populations, not three cuts of one dataset, and the survey instruments are not published. They should be read as evidence that measurement is weak, not as measurements. The consistent finding across all three is that the organizations reporting are largely unable to attribute outcomes. [19] [20]

ASSESSMENT The naive baseline, asking process participants how long their step takes and multiplying by a loaded hourly rate, fails in four specific ways. Self-reported durations are systematically inflated. Elapsed time and touch time are different quantities and get conflated. The rework loop, where a case comes back for correction, is usually excluded because the participants do not count it as part of the process. And the counterfactual is missing: volumes, staffing, and system changes over the measurement period get credited to the automation.

ASSESSMENT The disciplined version separates four measurements and captures each from a different source. Touch time, the sum of active human minutes, comes from task mining or timed observation rather than from recall. Elapsed time, from case creation to closure, comes from system timestamps. Rework rate, the share of cases that re-enter the process, comes from case history. And exception rate, the share of cases that leave the happy path, comes from the same logs. Automating a process without knowing its exception rate is how a program ships an automation that covers 60 percent of volume and reports 100 percent of the saving.

Step	What you capture	Where it comes from	What breaks if you skip it
1. Scope the case	Definition of one unit of work, start event, end event	Process owner sign-off	Every later number measures a different thing each time it is taken
2. Touch time	Active human minutes per role, per case	Task mining or timed observation, not recall	Self-reported estimates inflate the baseline and the saving is unauditable

Step	What you capture	Where it comes from	What breaks if you skip it
3. Elapsed time	Case creation to closure, including queue wait	System timestamps in the source of record	Queue time gets credited to the automation that did not remove it
4. Rework and exception rate	Share of cases re-entering the process or leaving the happy path	Case history and status transitions	Coverage gets overstated and post-automation volume looks like a saving
5. Cost basis	Loaded rate per role, agreed with finance before the build	Finance, in writing	The savings number is contested at the exact moment you need it
6. Holdout or staged rollout	A comparable unfulfilled cohort or a pre-and-post window with volume normalization	Deployment plan	External change gets attributed to the automation and the ledger loses credibility
7. Per-run telemetry	Duration, exception, human touches, model and infrastructure cost per instance	Control plane trace store	Return is recalculated annually from memory rather than measured continuously
8. Verified savings ledger	Claimed versus finance-accepted saving, per automation, per quarter	Control plane, reviewed by finance	The dashboard reports a number nobody outside the automation team believes

ASSESSMENT Two additions do most of the work and are almost always missing. The first is the distinction between claimed and verified saving, held as two separate fields in the ledger, with finance owning the second. A dashboard that shows only one number gets discounted to zero by the CFO the first time it is challenged. A dashboard that shows both, and shows the gap narrowing, becomes the instrument that funds the next tranche.

ASSESSMENT The second is capturing cost per run alongside time saved. An automation that removes eleven minutes of human effort and consumes four dollars of model inference per case has a different economic profile at ten thousand cases a month than at fifty, and only per-run telemetry surfaces that before the invoice does. Time-saved-only dashboards were adequate for deterministic RPA, where marginal execution cost was near zero. They are not adequate once model calls are in the path.

FACT Process and task mining tools exist to do the capture in steps two through four, and the distinction between them matters: process mining reconstructs the path from transactional system data, while task mining records what users actually do on the desktop. Celonis positions both, with task mining enriching the process model with desktop-level actions. [36]

ASSESSMENT Buying process mining to establish a baseline is defensible even in a program otherwise committed to owning its stack. It is a measurement instrument used at the start of an automation's life and periodically thereafter, not a runtime dependency, so the lock-in profile is mild and the alternative, hand-instrumenting every source system, is slow enough to delay the program past the point where anyone still cares.

ASSESSMENT One prediction worth holding the author to. Forrester expects process intelligence to rescue 30 percent of failed AI projects in 2026. The mechanism they describe, giving agents operational grounding, is plausible; the number is a forecast from an unpublished model and should be read as a claim that measurement is where recovery happens rather than as a rate anyone can verify. [13]

Audit-ready means an outsider can reproduce it without you

Audit readiness is not a document set. It is the property that a person who does not work for you, given read access, can independently establish what ran, on whose authority, against which version of the logic, and what it changed. Most internally built automation platforms fail on the third and fourth of those.

ASSESSMENT Four questions decide the outcome and they are the same in every framework. Who authorized this automation to exist and who owns it today. Which version of the code, configuration, and prompt was live at the time of the run in question. Which identity executed it and what could that identity reach. And what did it change, in a log the automation itself cannot alter.

FACT The control expectations for automated processes in a financially material path are established, not novel. Practitioner guidance in the SOX context has for years called for designed security, change management, and IT operational controls over bots, and flags the specific failure of consolidating several human steps into one automated process and thereby destroying a segregation-of-duties boundary that a control depended on. [32]

ASSESSMENT That failure mode deserves emphasis because AI-assisted construction makes it more likely, not less. The natural output of asking a model to automate a process end to end is a single identity performing initiation, approval, and posting. The human process had those steps separated because a control required it, and the separation is rarely documented in a form the model can see. Segregation of duties has to be an explicit design input to every automation, checked at review, or it will be silently removed.

Control objective	Artifact an auditor will accept	System of record
Authorization to operate	Signed request naming a business owner, data classes touched, and a criticality rating	Automation registry
Change management	Immutable link from each run to the commit, container digest, and prompt version live at execution	Trace store plus version control
Logic version control	Prompts and configuration versioned and promoted through environments, not edited in place	Prompt registry
Access and identity	One distinct non-human identity per automation, with owner, scope, expiry, and rotation evidence	Identity provider plus registry

Control objective	Artifact an auditor will accept	System of record
Segregation of duties	Documented mapping from human control points to automated steps, reviewed at design and at change	Design record, reviewed by internal audit
Execution evidence	Append-only run log with inputs, outputs, tool calls, exceptions, and cost, retained per policy	Trace store
Human oversight	Recorded approval events with the identity, timestamp, and the state presented to the approver	Trace store
Outcome measurement	Baseline, per-run telemetry, and finance-accepted savings, all three retained	Control plane ledger

FACT The regulatory position in the EU moved twice in 2026 and in opposite directions. Article 50 transparency obligations became generally applicable and enforceable by national authorities on 2 August 2026, covering systems that interact with people or generate synthetic audio, image, video, or text, regardless of whether the system is classified high-risk. Separately, the adopted AI Omnibus deferred the high-risk obligations to 2 December 2027, with a four-month extension to 2 December 2026 for machine-readable marking of synthetic content generated by systems already on the market before August. [21] [22] [23]

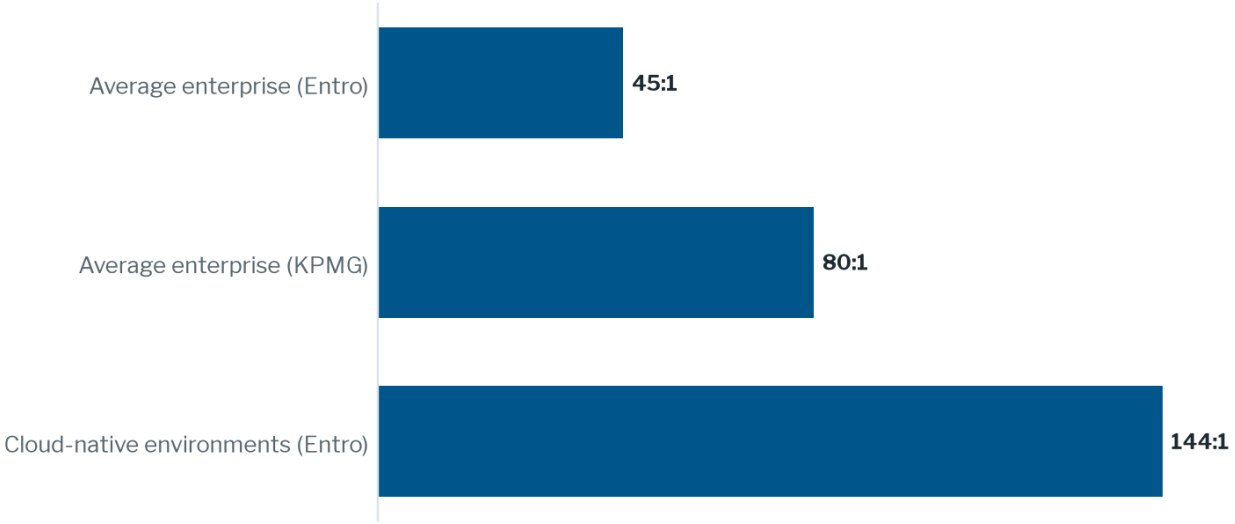
ASSESSMENT Deferral is the least useful thing to plan around. The obligations were not withdrawn, the evidence they require is the same evidence a SOX-scoped ITGC review already wants, and the cheapest time to build a trace store with retained inputs and outputs is before there are two hundred automations in production. Treating December 2027 as the deadline is how a team ends up retrofitting instrumentation into a live estate.

FACT ISO/IEC 42001 is now the certifiable management-system standard for AI, with certification bodies operational and major software vendors certifying against it; Pegasystems announced certification for its cloud and generative AI services on 3 February 2026, and Microsoft publishes its own 42001 position for its cloud services. [37] [38] [39]

ASSESSMENT For a buyer, 42001 is most useful as a procurement filter and least useful as an internal goal in year one. Requiring it of vendors costs nothing and narrows the field usefully. Pursuing it internally before the control plane produces evidence automatically means paying auditors to review a process that is still manual, which is the expensive order to do it in.

THIRD-PARTY ESTIMATE Identity is where the gap is widest. Vendor-sponsored counts put non-human identities at roughly 45 to 1 against human ones in the average enterprise, at more than 80 to 1 in KPMG's figure, and at 144 to 1 in cloud-native environments, up from 92 to 1 eighteen months earlier. The definitions of an identity differ between these counts and none of the underlying instruments are published, so the ratios are indicative. [29] [30]

FIGURE 7 Reported non-human to human identity ratios



THIRD-PARTY ESTIMATE Three separately sourced counts, reported second-hand through security trade press rather than from published methodology. Each uses a different definition of what counts as an identity, and both sources sell into the market the finding supports. The ratios should be read as evidence that the population is large and growing, not as measurements comparable to each other. [29] [30]

FIGURE 8 The agent governance gap, self-reported (percent of organizations)

THIRD-PARTY ESTIMATE Reported second-hand through security industry press; the original survey instrument, sample frame, and question wording are not published, and the two 92 percent figures may come from the same respondent pool asked adjacent questions. The gap between stated priority and implemented policy is the durable finding. [30]

ASSESSMENT The control that closes most of that gap is unglamorous and cheap: one identity per automation, never shared, owned by a named human, scoped to the minimum data classes, with an expiry date that forces a review. Doing this from the first automation costs nothing. Retrofitting it across two hundred automations that share three service accounts is a project, and it is the project that gets discovered during an audit rather than planned.

Failure modes, ranked by what actually ends the program

Ordered by how often each one terminates funding rather than by how loudly it is complained about.

#	Failure mode	How it ends the program
1	No credible baseline	The saving is asserted rather than measured. Finance discounts it to zero at the first budget challenge and the program loses its funding argument permanently. This is the most common cause and the cheapest to prevent.
2	Key-person concentration	Two engineers understand the control plane. One leaves. Velocity collapses and the organization discovers it bought a dependency rather than a capability. Owned stacks fail here far more often than they fail technically.
3	Maintenance backlog compounding	Year one ships forty automations. Year two spends most of its capacity keeping them running, and the roadmap stalls with nothing visible to show. AI-assisted authoring accelerates arrival at this point rather than delaying it.
4	Segregation of duties silently removed	An end-to-end automation collapses initiation, approval, and posting into one identity. It surfaces as an audit finding with a remediation deadline, and remediation means redesigning automations already in production.
5	Shared service accounts	One credential runs many automations. When something goes wrong, attribution is impossible and the blast radius is the union of every permission any automation needed.
6	Prompt drift without versioning	A prompt is edited in place to fix one case. Behavior changes for every other case. Without a version tied to each run there is no way to establish what the automation was doing last quarter.
7	Unmetered model spend	Inference cost per run is invisible until the invoice. An automation that was economic at pilot volume is not at production volume, and nobody notices until finance does.
8	Governance tooling treated as governance	An observability platform is installed and declared the control layer. It records; it does not authorize, meter, or enforce. The gap is found during the first audit.

ASSESSMENT The first two account for most terminated programs in the author's experience, and neither is a technology problem. That is the case for keeping the owned surface small: a control plane that four people can hold in their heads survives one resignation, while an estate of hand-built connectors does not.

THIRD-PARTY ESTIMATE The historical base rate is worth stating with its provenance attached. EY, one of the largest RPA implementation consultancies and therefore an interested party, reported that as many as 30 to 50 percent of initial RPA projects failed, and attributed the failures to methodology rather than technology. The figure dates to 2016, is repeated widely without its

date, and should be treated as an order-of-magnitude signal about programs rather than a current rate. [31]

Build versus buy, decided per layer

The recommendation is a set of layer-level decisions, not a single verdict, and the reasoning differs at each layer.

Layer	Decision	Reasoning
Screen-driven legacy RPA	Buy	The maintenance burden is per-application and never ends. This is the one place per-bot pricing is worth paying, because you are buying somebody else's obligation to keep up with UI changes in software you do not control.
Connector library	Buy, then selectively build	Several hundred maintained connectors is a multi-year commitment. Build only for the systems where your requirements are genuinely unusual, which is typically internal and industry-specific systems the vendor does not cover anyway.
Durable execution	Own	The open options are mature and the operational surface is well understood. Temporal, Restate, and Inngest all offer self-hosted paths, and this layer changes slowly, which is the profile that rewards ownership.
Workflow authoring	Either	Decide on who is authoring. If business analysts build, a visual tool earns its price. If engineers build, code and version control win on review, testing, and diff quality.
Agent runtime	Own	Framework churn is high and vendor abstractions here are thin. The lock-in cost of a proprietary agent runtime is high relative to what it saves.
Governance and control plane	Own	Your requirements diverge most from every other buyer's here, the switching cost of getting it wrong is highest, and it is the layer that produces the evidence you are ultimately accountable for.
Process baseline capture	Buy	A measurement instrument used at the start and periodically after, not a runtime dependency. Mild lock-in and the hand-built alternative delays the program badly.
Identity for automations	Buy, or extend what you own	Use the identity provider you already run. Building a second identity system for non-human principals is how organizations end up with two access reviews and one of them going stale.

ASSESSMENT The pattern in that table is consistent. Own the parts where your requirements are unusual and the change rate is low. Rent the parts where your requirements are ordinary and somebody else absorbs a maintenance obligation that never ends. Under that rule the per-bot execution license looks like the strongest candidate for cancellation and the connector library looks like the weakest, which is close to the opposite of how most of these decisions get made.

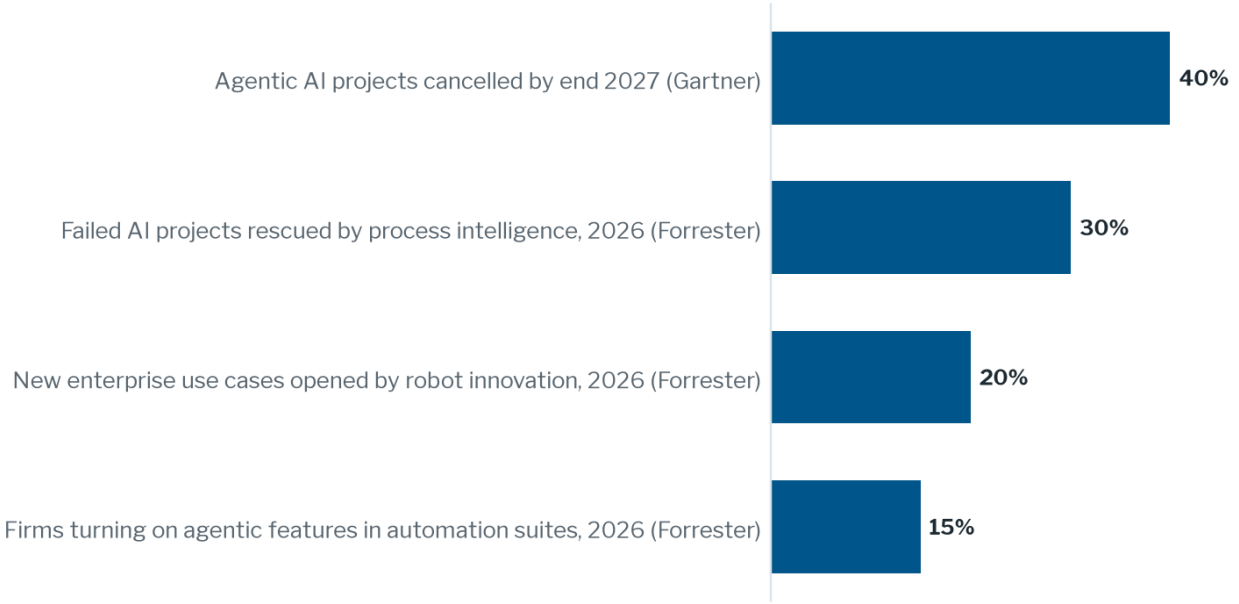
ASSESSMENT One sequencing recommendation carries most of the value. Build the control plane first, against the automations you already run on the incumbent platform, and only then start moving execution. A control plane that governs a rented estate is immediately useful, immediately auditable, and reversible. Migrating execution first produces a period, usually a year, in which nothing is governed and nothing is measured, and that period is where programs lose their sponsor.

Why point forecasts fail here, and what to watch instead

The category has repriced itself twice in three years. Any single-number forecast to 2028 is a guess wearing a decimal point, and the published analyst numbers illustrate the problem rather than solving it.

THIRD-PARTY ESTIMATE Gartner predicted on 25 June 2025 that over 40 percent of agentic AI projects would be cancelled by the end of 2027, citing escalating cost, unclear business value, and inadequate risk controls, and estimated that of the thousands of vendors claiming agentic capability, roughly 130 were genuinely delivering it. Forrester predicted on 3 November 2025 that fewer than 15 percent of firms would turn on agentic features in their intelligent automation suites during 2026, that process intelligence would rescue 30 percent of failed AI projects, and that robot innovation would open 20 percent of new enterprise use cases. [12] [13]

FIGURE 9 **Four published analyst predictions, and four different populations (percent)**



THIRD-PARTY ESTIMATE These four numbers are not comparable. Each measures a different population over a different window, and none of the underlying models is published. They are charted together to make that visible, not to rank them. The one with the clearest observable resolution is the Forrester adoption figure, because feature activation is something a vendor could disclose and a buyer could check at renewal. [12] [13]

ASSESSMENT The agent-washing estimate is the more useful of the two Gartner claims, and it is also the softer one. Gartner does not publish how it counted 130, so the figure cannot be checked. What it supports is a procurement posture rather than a market view: assume the agentic label

on a renewal quote is marketing until the vendor demonstrates a run that planned, selected a tool, recovered from a failure, and produced a trace you can read.

SCENARIO A Consolidation 40 percent

The incumbent suites absorb the agent layer and price it into existing seats. Governance features arrive as platform capability rather than as separate purchases, and the standalone observability and gateway tools become features.

Consequence. Owned control planes get harder to justify to a CFO, because the marginal cost of the rented equivalent falls toward zero. Programs that built early keep the portability advantage and lose the cost argument.

Earliest visible sign. A major suite vendor bundles agent tracing and prompt versioning into an existing SKU at no incremental list price.

SCENARIO B Two-speed split 30 percent

Regulated and high-assurance estates move to owned control planes over rented execution, while everyone else stays fully rented. The market bifurcates on compliance posture rather than on company size.

Consequence. The reference architecture in this report becomes the standard pattern in banking, telecommunications, healthcare, and public sector, and stays a minority pattern elsewhere.

Earliest visible sign. Two or more Fortune 500 firms in regulated sectors publish reference architectures showing self-hosted governance over vendor execution.

SCENARIO C Open control plane wins 20 percent

OpenTelemetry GenAI conventions stabilize, an interoperable evidence format emerges, and swapping the observability and gateway layers becomes routine. Governance stops being a lock-in point.

Consequence. The build case strengthens sharply, because the largest hidden cost of owning the control plane, re-instrumentation and migration risk, mostly disappears.

Earliest visible sign. The gen_ai.* conventions reach stable status and two or more commercial vendors ship native import and export against them.

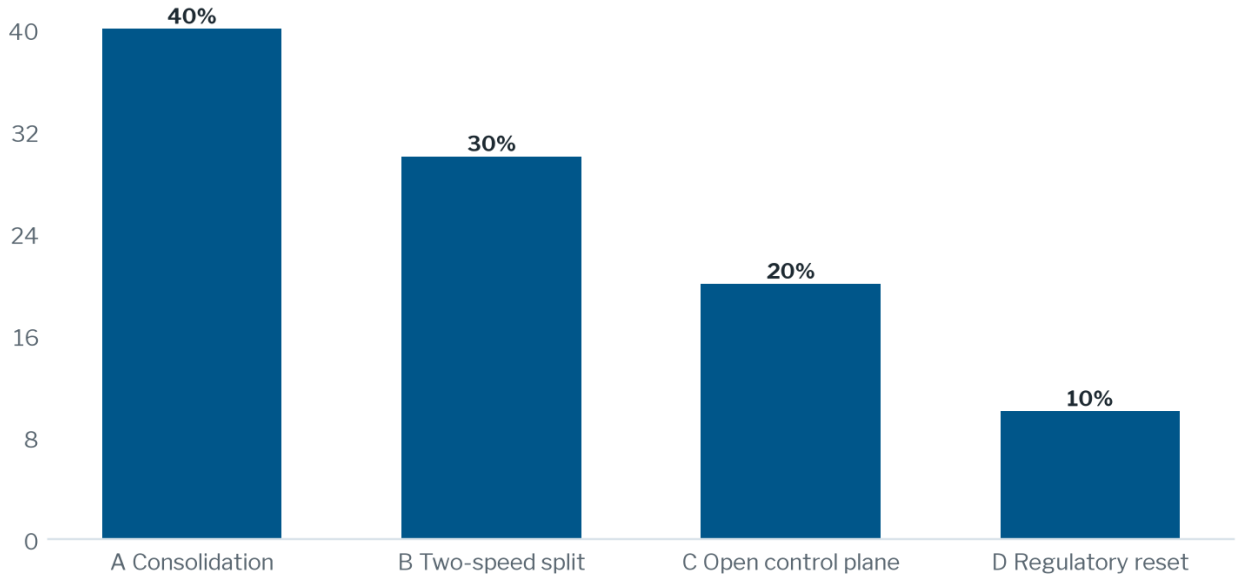
SCENARIO D Regulatory reset 10 percent

An enforcement action or a high-profile automated-decision failure moves liability onto deployers in a way that makes vendor attestations insufficient on their own.

Consequence. Evidence production becomes the first-order requirement and cost becomes secondary. Every layer decision gets re-made on auditability, and rented execution with no exportable trace becomes untenable.

> **Earliest visible sign.** A national regulator fines a deployer over an automated decision and names inadequate execution logging in the published decision.

FIGURE 10 **Scenario probabilities to end-2028, subjective**



SCENARIO The author's subjective probabilities, summing to 100. These are the least reliable content in this report and they are included so the reasoning can be argued with rather than because the numbers are defensible. The tripwires in the scenario table and the indicator table below are the useful part, because any reader can watch them without private information.

If this happens	It means	Moves toward	Where to watch
gen_ai.* conventions reach stable status	Trace portability becomes real and the migration risk of owning the control plane drops	Scenario C	OpenTelemetry GenAI conventions repository releases
A suite vendor bundles agent tracing into an existing SKU at no extra list price	The rented governance layer is being commoditized to defend seats	Scenario A	Vendor pricing pages and renewal quotes
UiPath net retention falls below 100 percent	The installed base is contracting and discounting will get steeper	Better buyer terms	Quarterly investor relations releases
ClickHouse gates a Langfuse core feature behind a commercial license	The open-core commitment is being tested and dependency risk is realized	Reassess trace store	Langfuse release notes and license file

If this happens	It means	Moves toward	Where to watch
A national regulator names execution logging in an enforcement decision	Evidence production becomes a first-order procurement requirement	Scenario D	Regulator decision publications
Two or more regulated Fortune 500 firms publish owned-control-plane architectures	The pattern has crossed from advocacy into reference practice	Scenario B	Conference talks and engineering blogs
Model API list prices rise for the tier you build on	The owned-stack cost model is understated in its largest variable line	Buy side of the ledger	Provider pricing pages
n8n or a comparable tool changes its self-hosting license	Relicensing risk in the owned stack is realized rather than theoretical	Reassess the stack	Repository license files and release notes

Implications

Four audiences, four different decisions. They are separated here because blending them produces advice that fits none of them.

For managers running an automation team

ASSESSMENT Your leverage point is the baseline, and it is available to you without any budget approval. Start capturing touch time, elapsed time, rework rate, and exception rate on the next three processes in your queue, before anyone commits to automating them. Agree the loaded cost rate with finance in writing first. Six months of that data changes what you can argue for more than any platform decision will.

ASSESSMENT Resist the pressure to ship automations faster because the model can write them faster. The constraint on your team is review and maintenance capacity, not authoring speed, and DORA's finding that AI increases delivery instability where control systems are weak describes exactly the position a small automation team is in. Add the review capacity before you add the throughput. [14]

ASSESSMENT Two habits cost nothing and prevent the two most expensive audit findings: one identity per automation from the first one, and prompts held in version control with a promotion path rather than edited in place. Both are trivial at automation number one and expensive at automation number one hundred.

For directors owning the platform decision

ASSESSMENT Do not present this as a license saving, because it is not one at mid-size scale and the model will not survive contact with finance. Present it as a control and portability decision with a cost that is roughly comparable to slightly higher, and put the break-even bot count in the paper so the conversation is about scale rather than about principle. [5]

ASSESSMENT Sequence the control plane before the execution migration. Build the registry, the trace store, the prompt registry, and the identity model against the automations you already run on the incumbent, and you get an auditable estate in months without a migration risk. This also gives you a working system to negotiate the renewal against, which is worth more than the engineering it costs.

ASSESSMENT Name the key-person risk in the paper before someone else does. An owned control plane needs four people who understand it, not two, and the second pair is the difference between a capability and a dependency. If the headcount is not fundable, the honest recommendation is to stay rented and spend the effort on the baseline instead.

For VPs setting the operating model

ASSESSMENT The center of excellence model most organizations run was designed to control citizen developers on a low-code platform. That is not the problem you have now. The problem is a small number of engineers producing a large volume of code quickly, which is a platform engineering problem with a different control set: code review, test coverage, deployment gates, and evidence generation rather than maker training and app inventory.

FACT Microsoft's own trajectory illustrates the shift. The Power Platform Center of Excellence Starter Kit, the reference governance implementation for that model, is no longer actively maintained, with its capabilities absorbed into the Power Platform admin center and issues no longer reviewed. [27]

ASSESSMENT Fund the platform team as a product team with a roadmap and users, not as a shared service with a ticket queue. The difference shows up in whether the control plane gets adopted by the automation builders or routed around by them, and a control plane that gets routed around produces exactly the ungoverned estate it was built to prevent.

ASSESSMENT Set the vendor posture explicitly: agentic claims on a renewal quote are marketing until the vendor demonstrates a run that planned, selected a tool, recovered from a failure, and emitted a trace your team can read. Gartner's estimate that roughly 130 of thousands of vendors claiming agentic capability actually deliver it is unverifiable in its precision and directionally right in its warning. [12]

For the C-suite

ASSESSMENT The decision worth your attention is not which automation platform. It is whether the organization can produce, on demand and without an engineer writing a query, an account of what its automated systems did and what that was worth. Everything else in this report is downstream of that capability, and it is the one that regulators, auditors, and acquirers all converge on.

THIRD-PARTY ESTIMATE The measurement gap is the governance gap. Roughly a quarter of AI initiatives are reported as delivering expected return and 16 percent as scaled enterprise-wide, on self-reported executive surveys with unpublished instruments. Whatever the true rate, an

organization that cannot attribute outcomes cannot allocate capital against them, and that is a board-level condition rather than a technology one. [19] [20]

ASSESSMENT Fund the baseline discipline before the platform migration, and fund it as a standing capability rather than a project. It is cheap, it is unglamorous, it produces the number every other investment decision in this area depends on, and it is the only part of this report that pays back regardless of which scenario plays out.

ASSESSMENT On the license question specifically: negotiate rather than exit, and negotiate now. A vendor holding 107 percent net retention on 11 percent ARR growth is a vendor protecting renewals, and the credible threat of an owned control plane is worth more at the table than the migration is worth on the balance sheet. [1]

What this document cannot answer

These need primary research, an engagement, or access to data that is not public. They are listed because a report that does not name its gaps is asking to be trusted rather than checked.

What large enterprises actually pay. Every tier above the entry point at UiPath is quote-only, and negotiated discounts at the top of the market are not observable from outside. The break-even calculation in this report uses list prices, which overstates the saving from cancelling. Only your own contract can settle it.

Whether AI-assisted development changes the maintenance cost of automations specifically. The GitClear and DORA evidence is about general software engineering. Automations are shorter, more numerous, and more integration-bound than typical application code, and it is plausible the effect is larger or smaller in this class. Nobody has published the study.

The failure rate of internally built automation control planes. Vendors publish RPA program failure rates; nobody publishes the rate at which internally built platform efforts are abandoned, because the organizations that abandon them do not write it up. This is the single most decision-relevant unknown in the report.

How auditors will actually treat model-mediated decisions in a financially material path. The control expectations for deterministic bots are settled. Whether a non-deterministic step in a control path is acceptable with adequate logging, or requires a deterministic gate, varies by audit firm and is being decided case by case right now.

The real total cost of a self-hosted governance stack at scale. Published pricing covers the commercial alternatives. The infrastructure, on-call, and upgrade cost of running Langfuse, a gateway, and a durable execution cluster at enterprise volume is known only to the teams doing it, and they are not publishing it.

Whether the connector maintenance burden is genuinely as large as assumed here. The assumption that a hand-built connector library is a multi-year commitment is drawn from practice rather than from a published study, and a rigorous accounting of connector maintenance hours per integration per year would materially sharpen the build-versus-buy line.

Half-life of this assessment

Decaying within six months: every published price in this report, the OpenTelemetry GenAI stability status, the specific funding and valuation figures, and the analyst predictions for 2026, which resolve or expire at the end of the year. Check the vendor pricing pages and the conventions repository before reusing any number here.

Decaying within twelve months: the assessment that the standalone observability and gateway tools remain independent purchases, which Scenario A would overturn; the UiPath negotiating position, which moves with two more quarters of retention data; and the ClickHouse open-core commitment, which has not yet been tested by a hard product decision.

Decaying within twenty-four months: the EU regulatory position, which resolves at the December 2027 high-risk deadline; the ISO 42001 procurement calculus, which changes once certification is common enough to stop differentiating; and the specific tool recommendations in the reference stack table, given the churn rate in this category.

Architectural, and unlikely to move: that the license is small relative to engineering payroll in a mid-size program; that connector maintenance is a permanent obligation rather than a one-time build; that segregation of duties has to be an explicit design input to automated processes; that one identity per automation is cheap at the start and expensive to retrofit; and that a measured pre-automation baseline is the only thing that converts a claimed saving into an auditable one. These follow from the structure of the problem rather than from the state of the market.

Limitations

This report was written independently. No vendor named in it reviewed it, funded it, or was given advance sight of it, and the author holds no commercial relationship with any of them. The recommendation to own the governance layer is consistent with the author's practice, which is a bias to declare rather than a bias that has been eliminated.

The evidence base has three specific weaknesses. Most of the quantitative research on AI-assisted development is produced by companies selling adjacent products, which is why those claims are labeled as vendor claims rather than facts and why the report leans on direction rather than magnitude. The identity ratio figures and the governance gap percentages are second-hand, reported through trade press without published instruments, and are the weakest numbers in the document. And the analyst predictions are unfalsifiable in their methodology, since none of the underlying models is published.

The cost model is the author's own and is sensitive to one input above all others: loaded cost per engineer. At $250,000 the gap between the paths widens; at $120,000 it approaches parity. It also assumes list pricing, no negotiated discount, and no migration cost, all of which flatter the owned path. A reader with real contract numbers should rebuild it rather than cite it.

Geographic scope is a limitation. The regulatory analysis covers the EU position in detail because that is where the timetable is explicit and public. US federal AI rules, state-level requirements, and sector regulators are moving on their own schedules and are not covered here.

Finally, the scenario probabilities are subjective and unverifiable. They are stated so the reasoning can be argued with. The tripwires attached to each one are the part intended to be used.

Sources

[1] UiPath, Inc. "UiPath Reports Fourth Quarter and Full Year Fiscal 2026 Financial Results." Investor relations release, fiscal year ended 31 January 2026. *Vendor primary* ir.uipath.com

[2] UiPath, Inc. "UiPath Reports First Quarter Fiscal 2026 Financial Results." Investor relations release, quarter ended 30 April 2025. *Vendor primary* ir.uipath.com

[3] UiPath, Inc. "UiPath Reports Second Quarter Fiscal 2026 Financial Results." Investor relations release, quarter ended 31 July 2025. *Vendor primary* ir.uipath.com

[4] UiPath, Inc. "UiPath Reports Third Quarter Fiscal 2026 Financial Results." Investor relations release, quarter ended 31 October 2025. *Vendor primary* ir.uipath.com

[5] Microsoft. "Power Automate pricing." Published list prices, accessed August 2026. *Vendor primary* microsoft.com

[6] UiPath, Inc. "Pricing." Published plan tiers, accessed August 2026. *Vendor primary* uipath.com

[7] n8n. "Pricing." Published plan tiers and self-hosting terms, accessed August 2026. *Vendor primary* n8n.io

[8] Temporal Technologies. "Pricing." Published tiers and storage rates, accessed August 2026. *Vendor primary* temporal.io

[9] ClickHouse. "ClickHouse welcomes Langfuse: The future of open-source LLM observability." 16 January 2026. *Vendor primary* clickhouse.com

[10] ClickHouse. "ClickHouse raises $400M Series D led by Dragoneer to accelerate expansion across analytics and AI infrastructure." 16 January 2026. *Vendor primary* clickhouse.com

[11] Langfuse. "Self-host Langfuse." Product documentation, accessed August 2026. *Vendor primary* langfuse.com

[12] Gartner. "Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027." Press release, 25 June 2025. *Analyst firm* gartner.com

[13] Forrester. "Predictions 2026: Automation At The Crossroads." Blog preview of Predictions 2026: Automation And Robotics, 3 November 2025. *Analyst firm* forrester.com

[14] DORA and Google Cloud. "State of AI-assisted Software Development 2025." Annual research report. *Vendor primary* dora.dev

[15] DORA. "The ROI of AI-Assisted Software Development," in DORA Publications, 2026.01. *Vendor primary* dora.dev

[16] GitClear. "The Maintainability Gap: 2026 AI Code Quality Research." January 2026. GitClear sells code-quality analytics. *Vendor primary* gitclear.com

[17] Veracode. "2025 GenAI Code Security Report: Assessing the Security of Using LLMs for Coding." Veracode sells application security testing. *Vendor primary* veracode.com

[18] MIT Media Lab, Project NANDA. "The GenAI Divide: State of AI in Business 2025." Preliminary working paper, not peer reviewed; reported via press. Cited in this report only to explain why it is not relied upon. *Second-hand (unverified)* finance.yahoo.com

[19] IBM. "IBM Study: CEOs Double Down on AI While Navigating Enterprise Hurdles." Newsroom release, 6 May 2025, based on an IBM Institute for Business Value survey of 2,000 CEOs. *Vendor primary* newsroom.ibm.com

[20] IBM. "How to maximize AI ROI in 2026." IBM Think insights. *Vendor primary* ibm.com

[21] Goodwin Procter. "EU AI Act Transparency Obligations Are Now in Force." Client alert, August 2026. *Independent press* goodwinlaw.com

[22] Norton Rose Fulbright, Data Protection Report. "The EU AI Act: when does it become enforceable now?" July 2026. *Independent press* dataprotectionreport.com

[23] EU Artificial Intelligence Act. "Article 50: Transparency Obligations for Providers and Deployers of Certain AI Systems." *Standards body primary* artificialintelligenceact.eu

[24] National Security Agency. "Model Context Protocol (MCP)." Cybersecurity Information Sheet. *Standards body primary* nsa.gov

[25] OpenTelemetry. "Semantic conventions for generative AI." Dedicated GenAI conventions repository, split out at v1.42.0 on 12 June 2026. *Standards body primary* github.com

[26] OpenTelemetry. "GenAI semantic conventions" relocation notice, main specification site. *Standards body primary* opentelemetry.io

[27] Microsoft Learn. "Power Platform Center of Excellence (CoE) Starter Kit overview." Carries the notice that the kit is no longer actively maintained; page updated 22 June 2026. *Vendor primary* learn.microsoft.com

[28] Microsoft Learn. "Managed Environments overview." Power Platform administration documentation. *Vendor primary* learn.microsoft.com

[29] Cloud Security Alliance. "The Non-Human Identity Governance Vacuum." Research whitepaper on non-human identity and agentic AI governance. *Practitioner aggregate (self-selected)* cloudsecurityalliance.org

[30] Security Boulevard. "The Agent Identity Problem: Non-Human Identities Outnumber Humans 45 to 1 and AI Agents Are Making It Worse." July 2026. Aggregates Entro Security and KPMG figures; original instruments not published. *Second-hand (unverified)* securityboulevard.com

[31] EY. "Get ready for robots: Why planning makes the difference between success and disappointment." 2016. EY is a large RPA implementation consultancy. *Vendor primary* eyfs.ie

[32] Protiviti. "Building Bot Boundaries: RPA Controls in SOX Systems." The Protiviti View. *Practitioner review* blog.protiviti.com

[33] Temporal Technologies. "Temporal raises $300M to make agentic AI real for companies." 17 February 2026. *Vendor primary* temporal.io

[34] n8n. "n8n raises $180m to get AI closer to value with orchestration." Company blog, October 2025. *Vendor primary* blog.n8n.io

[35] Reuters, via Yahoo Finance. "Anthropic aims to nearly triple annualized revenue in 2026." Reported figures are sourced to unnamed people familiar with the matter and are not company-disclosed. *Second-hand (unverified)* finance.yahoo.com

[36] Celonis. "Process Mining" and "Task Mining" product documentation. *Vendor primary* celonis.com

[37] ISO. "ISO/IEC 42001 explained: what it is." Standards body resource. *Standards body primary* iso.org

[38] Pegasystems. "Pega Achieves ISO/IEC 42001:2023 Certification." Business Wire release, 3 February 2026. *Vendor primary* businesswire.com

[39] Microsoft Learn. "ISO/IEC 42001:2023 Artificial Intelligence Management System Standards." Microsoft compliance offering. *Vendor primary* learn.microsoft.com

Rights and permitted use

This document is analysis, not investment, legal, or procurement advice. It was not commissioned, reviewed, or funded by any company named in it, and the author holds no position in any of them.

Market Assessment reports, commissioned research, and briefings: **davidsoden.com/market-assessment**