



SKYCATCHER RESEARCH

Machine Hour

Rise of the Machine Economy · the thesis, edition of October 8, 2026

PDF regenerated October 23, 2026 · source dates shown with each series.

Contents. Each entry links to its page; the sidebar outline carries every section.

Research commentary from Skycatcher, LLC, an exempt reporting adviser, not investment advice. [Disclosures: Appendix D3, page 147.](#)

Executive summary

p. 2

The thesis in brief: more machine work, the infrastructure it needs, and the evidence that would prove us right or wrong.

I A new source of work

p. 12

Why agents can expand the amount of work software performs.

A new source of work · Cheaper work expands what gets attempted · Count the time agents actually work · The evidence is strongest in software · Two paths to 2032 · The prediction we are making

II Counting machine hours

p. 23

How we estimate current activity and build the 2032 scenarios.

Counting machine hours · What the public data can tell us · From today's estimates to the scenarios · How many people use agents · How long each agent works · Capability and working time are different · Our own measurements · The activity behind the OpenRouter estimate · From hours to tokens · Compute can limit the outcome · How we will judge the crossing

III The infrastructure behind the work

p. 34

How agent activity reaches infrastructure and when that use becomes revenue.

The infrastructure behind the work · Five layers of the system · Who serves the new demand · Four ways an hour reaches a bill · What the companies have reported · From working time to revenue · The opportunity beyond current software budgets

IV How we will test the thesis

p. 52

Financial scenarios, market expectations and the evidence that would change our view.

How we will test the thesis · What has the market rewarded so far? · What do today's prices assume? · What must the businesses deliver? · Can activity grow without much provider revenue? · What would change our view · What evidence will settle the argument? · How does the test differ by company?

V Building for agents

p. 69

New companies, simpler backends and the longer-range vision of an operating layer for the machine economy.

Building for agents · Where new companies are forming · Funding the backend build-out · Fewer systems to connect · What adoption can establish · From assembling services to running software · What we have learned by building · An operating layer for the machine economy · What would make that vision real

Appendices

p. 86

Sources, assumptions, definitions, company records and the original terms of our predictions.

Tokens: the history, our forecast and the outside view · The machine hours model: the assumption register and Jevons' paradox · The companies: the roster, the value chain and the private record · The record: downside, registered calls and disclosures · Company presentations: findings summarized in our own illustrations

About

p. 159

Who publishes this research, what we study, how to follow it, contact and disclosures.

What we observed · What we think it means · Who we are · How we work · Write to us · Disclosures

Tracker

p. 161

The latest published evidence on agent activity, backend revenue, prices and capability. Each series carries its own date and method.

Where things stand · The indices · Fundamentals · Valuation · Capability · Production · Tokens

SKYCATCHER RESEARCH · THE FULL THESIS

Machine Hour

Five parts follow the work, the measurement, the infrastructure, the evidence and the systems being built for agents.

PRINTED OCTOBER 23, 2026 FROM MACHINEHOUR.AI

Executive summary Part One — A new source of work Part Two — Counting machine hours
Part Three — The infrastructure behind the work Part Four — How we will test the thesis
Part Five — Building for agents Appendices — Evidence, methods and prediction records About

Executive summary

We see a future where billions of AI agents work around the clock. By the end of 2032, we expect this new workforce to exceed 7 trillion working hours a year, surpassing all paid human work in 2026. Our Machine Economy scenario reaches 340 trillion hours a year, 49 times that scale. This is a once-in-a-generation transformation in how humans work.

We came to this view by putting agents to work ourselves. They read filings, write software, check results and keep research current. A question can become an ongoing research process. A small team can build software, operate it and improve it as needs change. One person can direct many streams of work at once.

We measure this activity in machine hours. One agent actively working for one hour counts as one machine hour. Agents working in parallel count separately; idle time does not. This measures the scale of activity. Its value depends on the quality of the results, the cost of human review and whether people continue to use it.

Our two scenarios explore how far this could go. The **Agentic Workforce** reaches **19 trillion machine hours a year in 2032**. The **Machine Economy**, where agents do much more work without step-by-step direction, reaches **340 trillion—49 times our 2026 human-work benchmark**. Both require rapid adoption. The larger case also requires major gains in computing capacity, efficiency and reliability. These are scenarios to test as evidence develops.

As useful work becomes cheaper, more becomes worth attempting. Research can cover more ground. Small teams can run services that once required large organizations. People can pursue ideas that were previously too expensive to build or maintain.

Every one of those services needs somewhere to run. Agents need access to information, permission to act and a reliable record of what happened. Their software needs to keep working after the conversation ends. This is the opportunity we study in the backend: the infrastructure that runs software and preserves its work.

That opportunity spans cloud platforms, established providers, open-source projects and new companies. More activity creates demand, but competition, falling prices and customer choices will determine how much becomes lasting revenue.

The longer-term vision is an internet that works on our behalf. People set the goals and limits. Agents assemble the software, coordinate work across clouds and devices, and keep services running. Building the shared infrastructure that makes this dependable could expand what individuals and small teams can accomplish. **We see a future where far more of what people imagine becomes something they can put to work.**

[Skycatcher](#), October 8, 2026

In May 1995, Bill Gates described the shift ahead in [The Internet Tidal Wave](#). His memo is a useful reference for the question we face now: what changes when the Internet gains a new source of work?

The Internet is a tidal wave.

It changes the rules.

BILL GATES, INTERNAL MICROSOFT MEMO · MAY 26, 1995

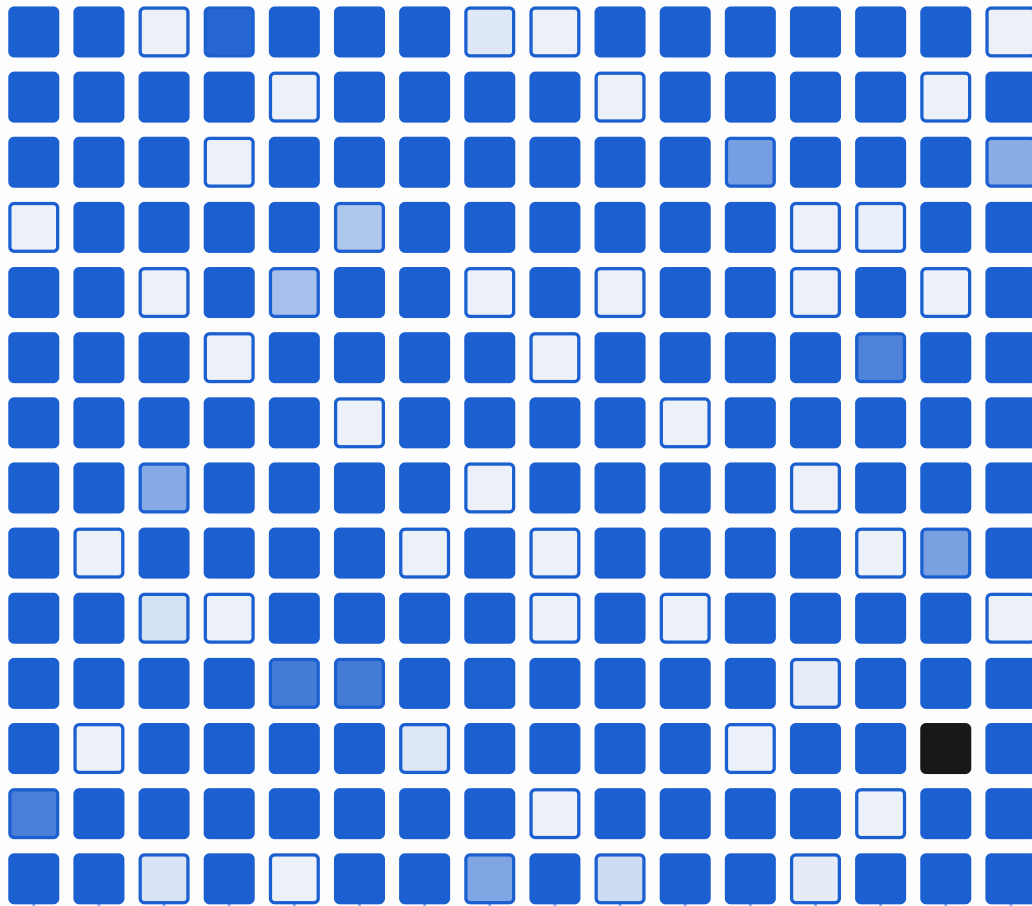
To	Executive Staff and direct reports
From	Bill Gates
Date	May 26, 1995

The Internet Tidal Wave

The memo's header, from Government Exhibit 20 in [United States v. Microsoft](#). From the Department of Justice record; reproduced for comment, and copyright remains with Microsoft.

We believe agents can change the scale of work the Internet performs. Machine hours give us a way to study that change.

AGENTS



THE BACKEND

THE BACKEND · STATE AND WORK · 32 PUBLIC COMPANIES

Work

executes, coordinates and accounts for the hour

State

keeps the record and the memory

Intelligence

the models that reason over the hour's data and answer

Computation

the chips and the energy beneath it all

■ at work

□ waiting

● one hour's round trip, traced every ten seconds

Agents use backend services, which call models running on computing hardware. Results return through the system and become the next step's input.

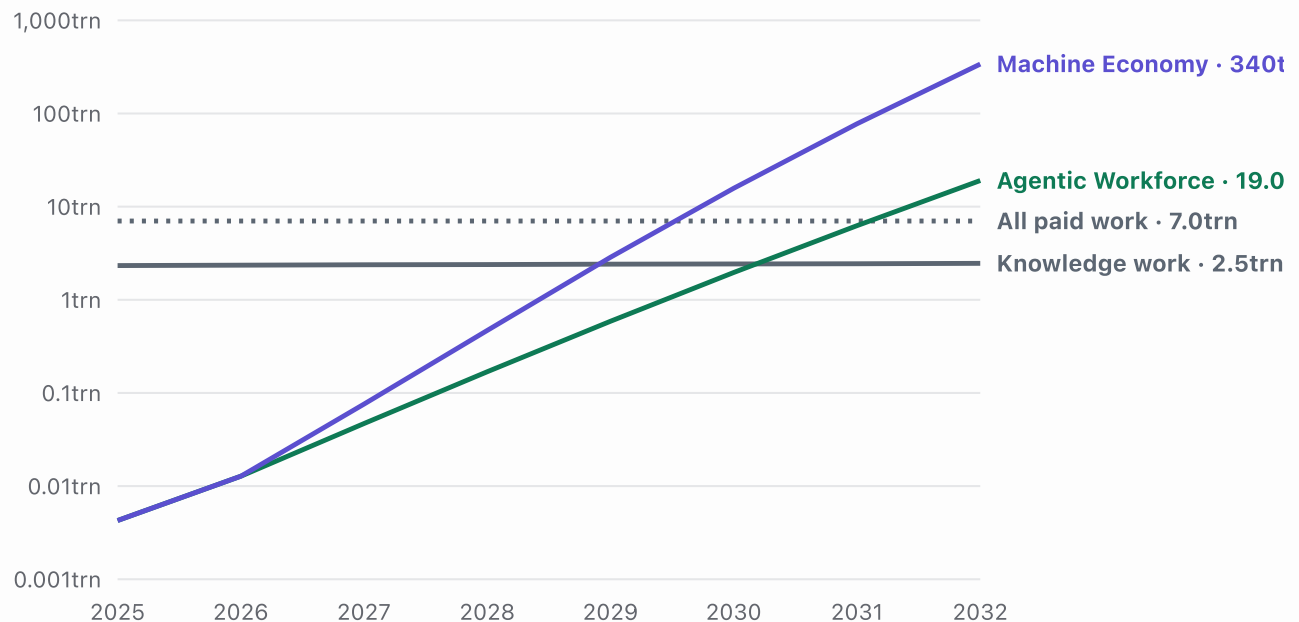
Figure 0.1 sets the scale. Both lines are modeled annual rates of machine hours against all paid human work; the parts that follow carry the evidence behind them, and the model lets a reader draw a different case.

FIGURE 0.1

SCENARIO

Machine hours a year: two scenarios against all paid human work

Machine hours a year against all paid human work and human knowledge work, log scale; move the two sliders to draw your own case

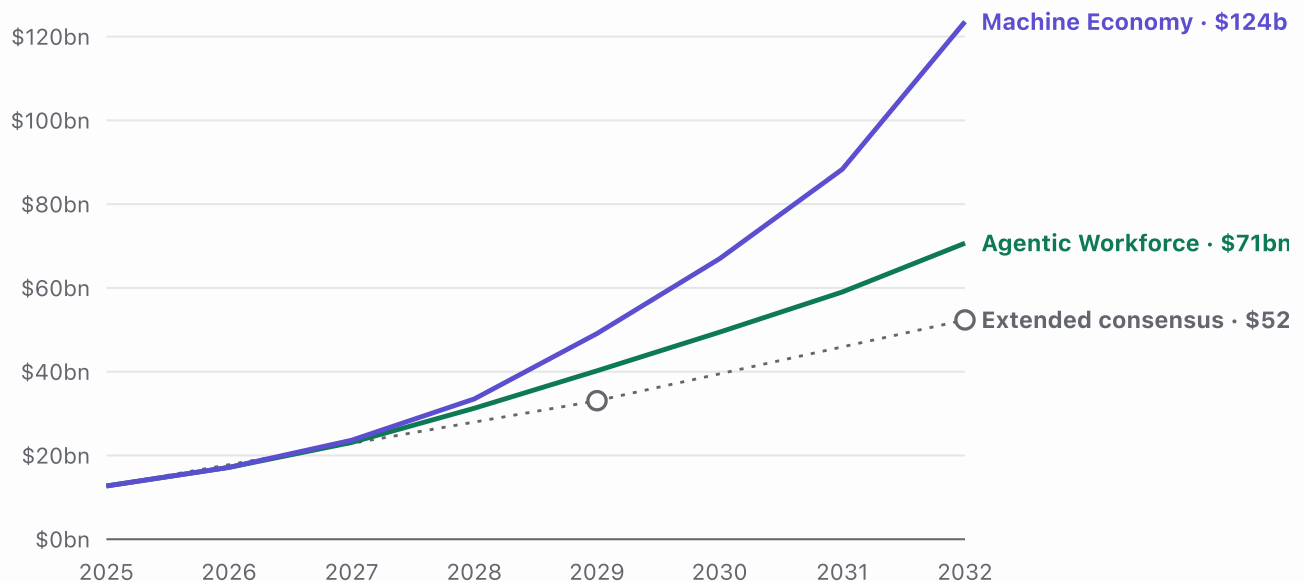


[Open the full machine hours model →](#)

Modeled annual rates on a logarithmic scale. The Agentic Workforce reaches 19 trillion hours in 2032 and the Machine Economy 340 trillion, on the stated capacity assumptions. These are scenarios, not bounds. The human lines compare time, not productivity or jobs displaced. The model lets a reader draw their own case. Passing all paid human work requires about 2.86x a year from the 2026 anchor for six years; the Agentic Workforce path grows 3.38x a year and the Machine Economy 5.46x. [Figure 1.9](#) carries the test.

Revenue at four companies under two scenarios

Revenue for the four company cases, \$bn a year: both scenarios on their own hours, the consensus case ringed in 2029 and 2032; the sliders price your own hours on the Machine Economy's inputs



[Open the revenue model behind this figure →](#)

The four-company scenarios reach \$71 billion and \$124 billion of annual revenue in 2032. Most of the near-term difference from consensus comes from our existing-business growth assumptions. The comparison uses analyst consensus through 2029 and our extension after that, with growth fading to 12% by 2032. The [model](#) also includes slower adoption at \$47 billion and limited provider revenue at \$59 billion in 2032. All four are available in the [scenario controls](#). About nine tenths of the 2029 revenue above consensus comes from the existing-business growth assumption. Agent revenue matters more later and follows delivered hours. The 340-trillion-hour case needs nearly all the capacity in the post-2029 build; Appendix A3 sets out those assumptions.

More work becomes worth attempting

Agents can choose steps, use tools and check a result. For selected coding tasks, we estimate about \$1 of model use against roughly \$200 of professional time, before review and other costs. Review, integration and operating costs remain. The larger effect is what becomes worth attempting: a report becomes a monitor, or a checklist becomes a process that runs. [Part One](#) explains the demand case.

Counting the work

Active agents multiplied by hours per agent gives machine hours. In our scenarios, the average reaches about three hours a day in the Agentic Workforce and 13.6 in the Machine Economy. The difference rests mainly on autonomous work, which remains the least proven assumption. Machine hours pass the seven-trillion-hour human benchmark in 2032 and 2030 respectively. [Part Two](#) opens the calculation.

Where the work runs

We map the stack of machine labor in five layers, from hardware to applications. The backend is State (data and memory) and Work (execution and control). Agents need both. Our public-company universe follows thirty-two providers, while the wider research includes clouds, private companies and open-source alternatives. Competition, pricing and customer choices determine who earns lasting revenue.

Cloudflare, MongoDB, Snowflake and Datadog give us four detailed examples of how work reaches a bill. Their combined revenue growth rose from about 25% in the first quarter of 2025 to about 34% on the latest reported quarters, to July 31, 2026, the baseline for bet B11. Their AI disclosures offer further evidence, with different definitions. The timing does not tell us how much growth came from agents. [Part Three](#) follows the activity into the companies' reported results.

Market expectations: a separate test

The research baskets show the hardware build being rewarded ahead of much of the backend. At September 30, 2026, Computation was up about twelvefold from January 2022. The backend basket, which includes Palantir, was up about 99%, against 84% for the Nasdaq 100; Palantir accounts for much of that gain. These are hypothetical baskets selected in 2026 and applied to earlier prices. They provide context for expectations, not a record of investment returns.

We believe the order of events is what matters. Through 2029, our financial scenarios depend mostly on stronger growth in the existing businesses. Later, agents become a larger source of revenue. From \$12.7 billion in 2025, the four-company scenarios reach \$71 billion and \$124 billion in 2032, against about \$52 billion in the extended consensus case shown in Figure 0.2. These outcomes depend on capacity, useful demand, provider share and pricing. More activity can coexist with lower provider revenue if customers choose other systems or prices fall faster. [Part Four](#) tests the financial paths and sets out what would change our view.

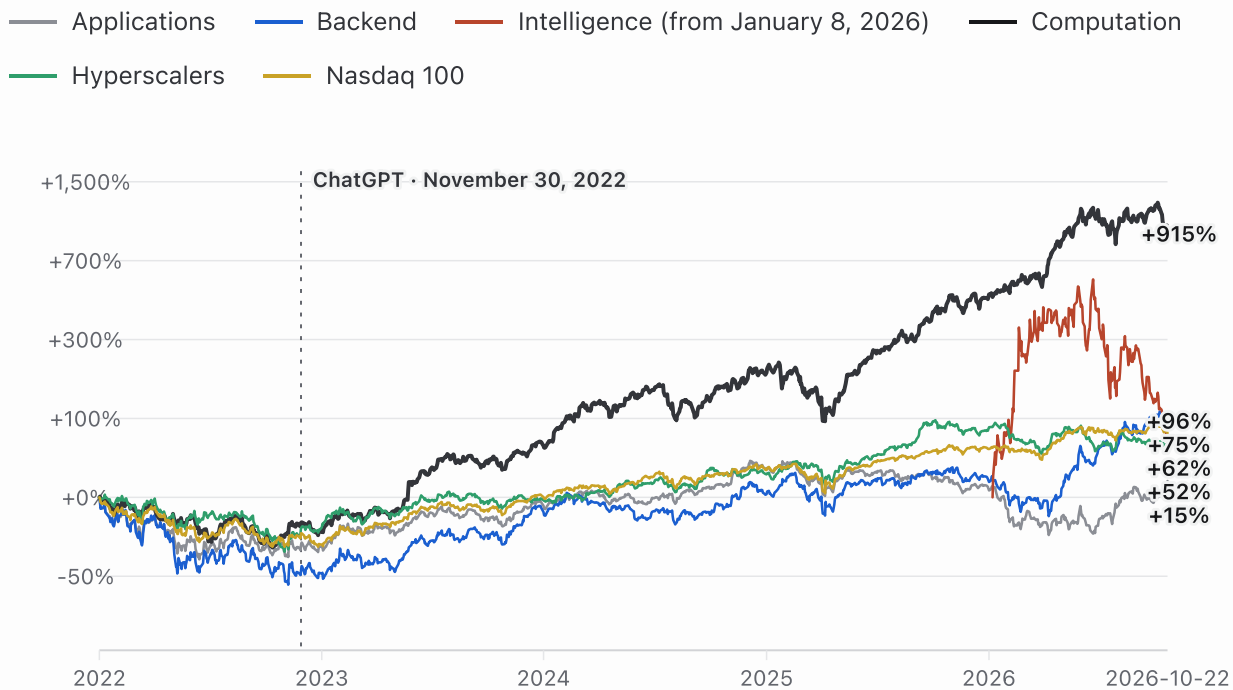
FIGURE 0.3

HYPOTHETICAL RESEARCH BASKETS · HISTORICAL PRICES, UPDATED DAILY

The layers of the stack since January 3, 2022

Pick the start date and the layers; every line is rebased to 0% at the start

EQUAL-WEIGHT INDICES · CHANGE SINCE JANUARY 3, 2022 · LOG SCALE · THROUGH OCTOBER 22, 2026



[Follow the layers live on the tracker →](#)

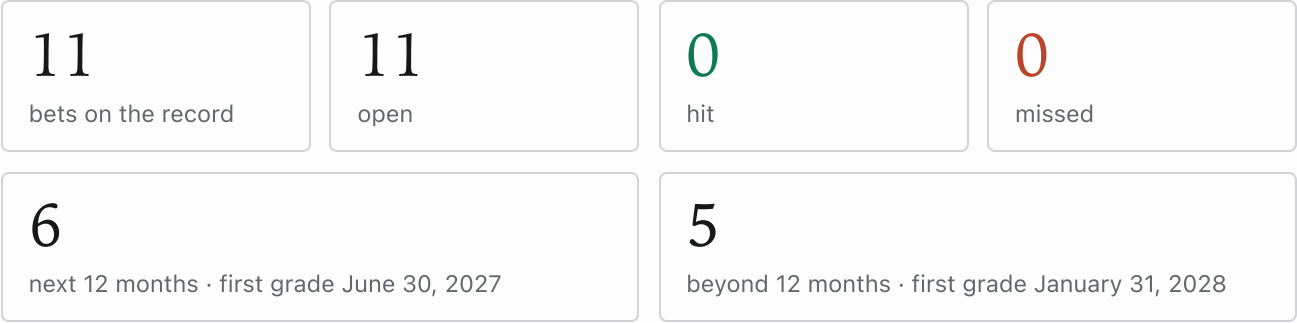
Daily price record. The equal-weight baskets were defined in 2026 and applied back to 2022, introducing hindsight. They are hypothetical and exclude fees and costs. This is also Figure 4.1 and the [tracker's index chart](#).

Our dated market predictions remain on the record alongside the operating tests. From October 1, 2026, we expect the backend basket to outperform Computation over one year and lead the listed comparison groups over three. The tracker grades the result from that start date.

FIGURE 0.4

OBSERVED · DATED RECORD

Eleven bold bets on the record



Eleven supporting predictions, each with a measure and deadline, including an earlier machine-hour milestone for the end of 2027. The first grades fall in 2027. We publish misses alongside successes. See the [tracker](#), Figure 4.8 and the full terms in Appendix D2.

In a personal post on August 19, 2026, Jerry Tworek, formerly a research vice president at OpenAI, reduced the future of computing to databases, neural networks and chips. It is a useful way to frame the question: as models do more work, what systems must keep that work running and preserve its results?



Jerry Tworek ✓

@MillionInt



What will be left at the end of computer science is databases, neural networks and chips

1:51 PM · Aug 19, 2026 · **203.2K** Views



123



133



2K



330



Elon Musk ✓



@elonmusk · Aug 19



Sounds probably right



161



90



1.3K



103K



Jerry Tworek on X, August 19, 2026, and Elon Musk's reply. Screenshot of personal posts, reproduced for comment; copyright remains with the authors.

We believe machine work can become one of the largest new sources of software demand since the web. **Machine Hour** follows that possibility from the work agents perform to the infrastructure it requires. The thesis states our view, the model shows what it takes, and the tracker tells us where the evidence leads.

This website is research commentary from [Skycatcher](#), published to share an idea for general research and discussion: not an offer, a recommendation or investment advice, it does not consider individual circumstances, and it is not intended for distribution where that would be unlawful. The indices are hypothetical and not investable, and the scenario figures are our estimates, not company guidance. [Appendix D3](#) carries the disclosures.

I

A new source of work

Why agents can expand the amount of work software performs.

9 MIN READ · 9 FIGURES

[A new source of work](#) [Cheaper work expands what gets attempted](#) [Count the time agents actually work](#)
[The evidence is strongest in software](#) [Two paths to 2032](#) [The prediction we are making](#)

A new source of work

Software has automated fixed processes for decades. Agents extend that ability: given a goal, they can choose steps, use tools, inspect the result and decide what to do next. A person can direct the task, or an event can start it under rules a person has set.

We call the resulting activity machine labor. It includes the attempts that fail. People remain responsible for the authority they give an agent and for deciding whether its work is good enough to use.

The change matters because software demand can grow beyond the number of people using it. One person can run several agents at once. A business can assign agents to processes that continue outside the working day. The limit becomes the supply of useful tasks, the cost of completing them and the infrastructure available to run them.

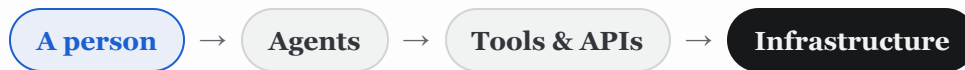
From a person's goal to machine work

The cloud era CLOUD-ERA PATTERN



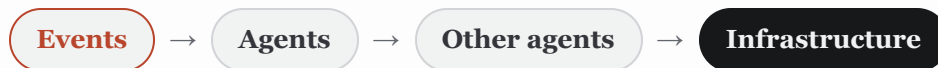
Applications and existing automation translate human goals into infrastructure demand.

The agent era AGENT PATTERN



A person sets the goal; agents do the work, each goal becomes many machine actions.

The machine economy CONDITIONAL DIRECTION



No person at the start of the chain, machines respond to machines, continuously.

A person can use software directly, delegate a task to an agent, or set rules that let an event start the work. Each path uses infrastructure. The diagram shows how an agent can add steps and call other agents before the task is complete.

Cheaper work expands what gets attempted

For selected coding tasks, we estimate about \$1 of model use against roughly \$200 of professional time, before review and other costs. The comparison excludes review, retries, integration and the cost of running the finished software. It illustrates the change in direct production cost; it is not a measured reduction in the full cost of software development.

The more interesting effect is what becomes worth building. A spreadsheet can become a live dashboard. A weekly report can become a monitor. A checklist can become a process that carries out its own steps. People who could describe a problem but could not write software can now attempt a solution.

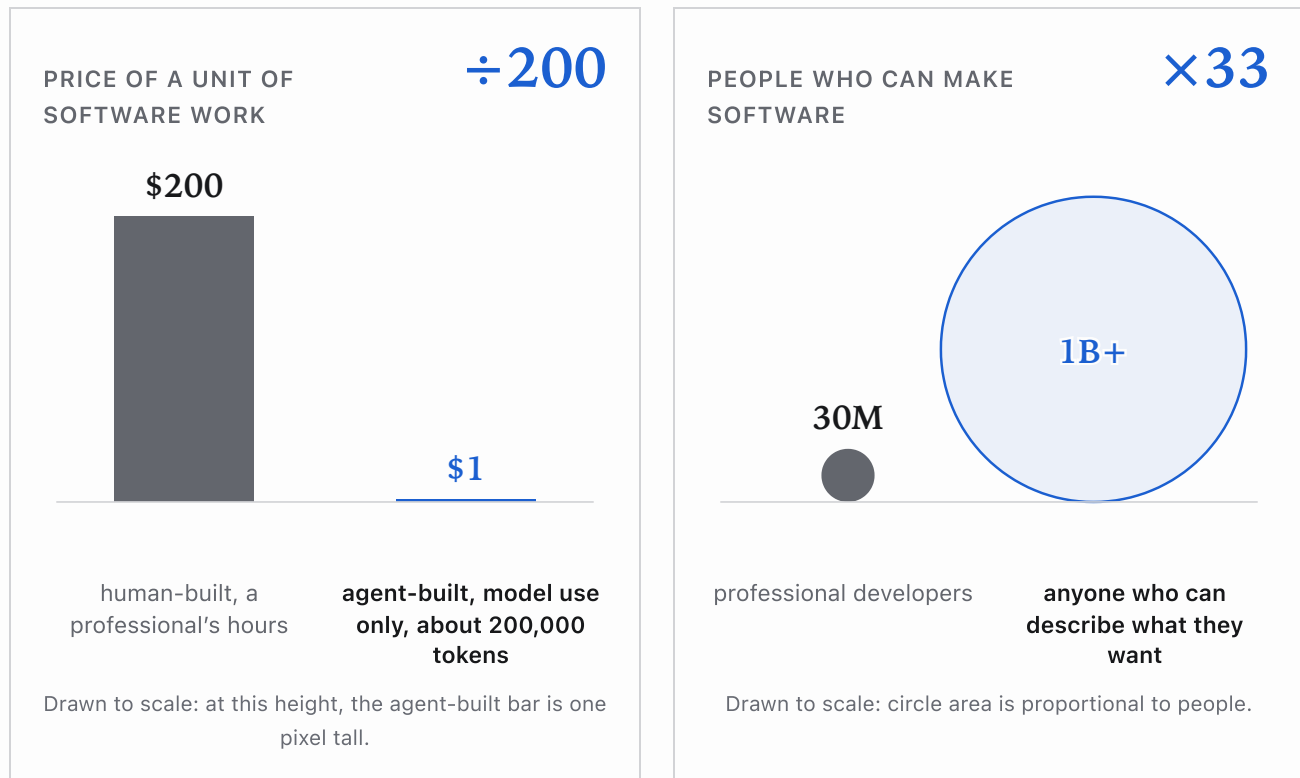
We expect much of the new work to come from tasks that were previously too expensive. An analyst can investigate more companies. A small business can test more ideas. A security team can examine more events. The outcome depends on whether the additional work is useful enough to justify its full cost.

Lower prices can expand use. Total spending rises only if the increase in use more than offsets the lower price. We think this is plausible in AI, but it will not happen automatically or equally in every market.

FIGURE 1.2

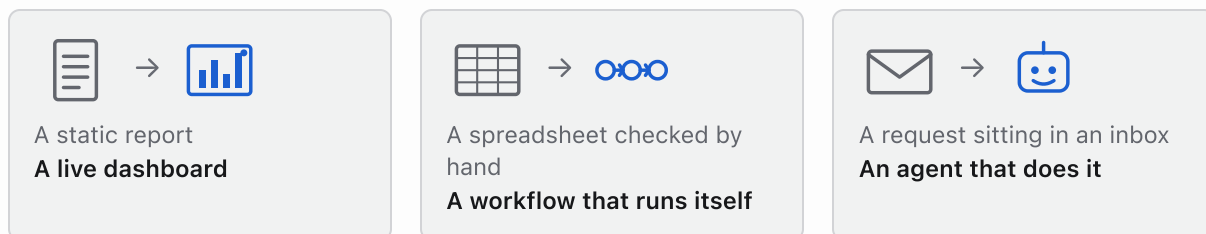
ILLUSTRATIVE

Lower costs make more software worth attempting



= More software worth building

Software and automation reach every corner of knowledge work, made by anyone who can describe the work



more work creates demand for computing, data and reliable execution

This illustration combines a 200-fold reduction in direct production cost with a 33-fold expansion in potential builders. Neither multiplier is a measured change in completed, useful software. The one-billion-person pool represents people who might build, not active developers or paying customers. Review, integration and operating costs still matter.

Count the time agents actually work

An agent that exists for a month may work for only a few minutes. An agent count therefore tells us little on its own.

A machine hour is one hour of active execution by one agent process. Ten agents working for an hour contribute ten machine hours. An agent waiting for an event contributes none. When a parent waits for a child agent, we count the child’s work without also counting the parent’s wait.

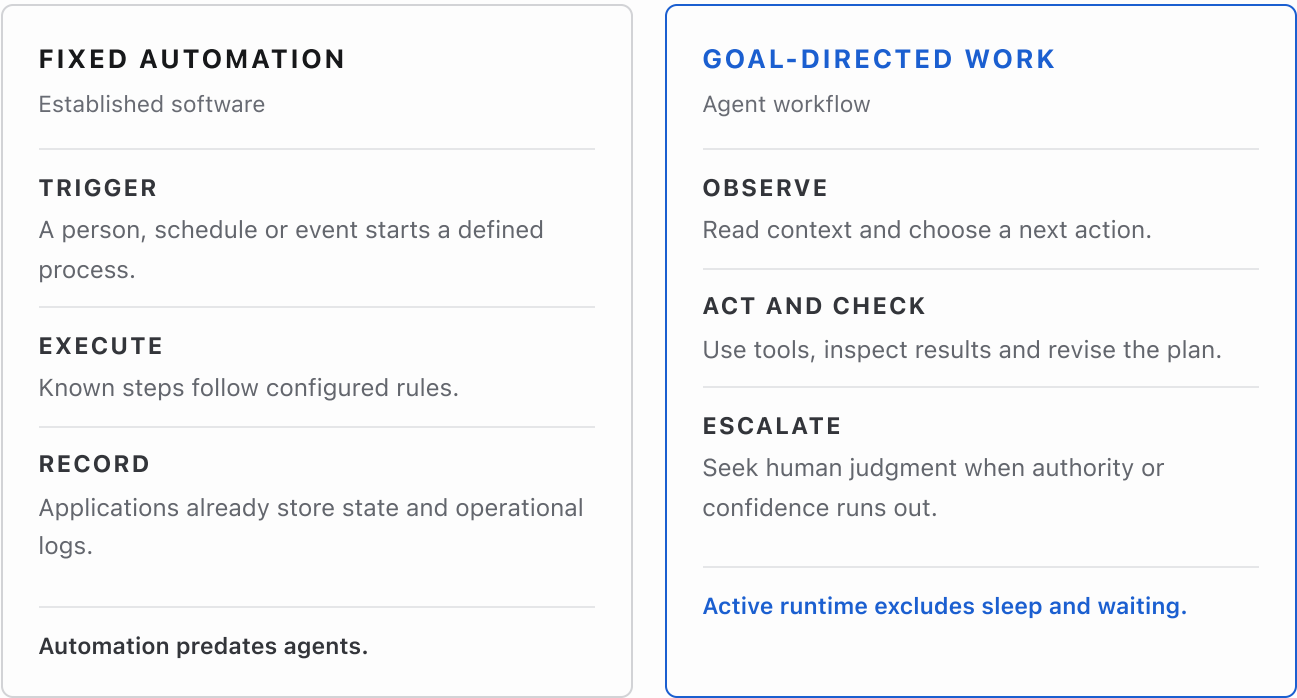
One instruction can produce many steps: retrieve information, call a model, query data, run a test, save a result and check it. The time spent on those steps is what we want to measure. A model call, a backend action and a completed task are different units.

Each hour also has a different infrastructure cost. Some agents run cheaply on a local machine. Others use a warehouse of business data or a large model. Part Three follows those differences into the bill.

FIGURE 1.3

ILLUSTRATIVE

One request can start a working loop



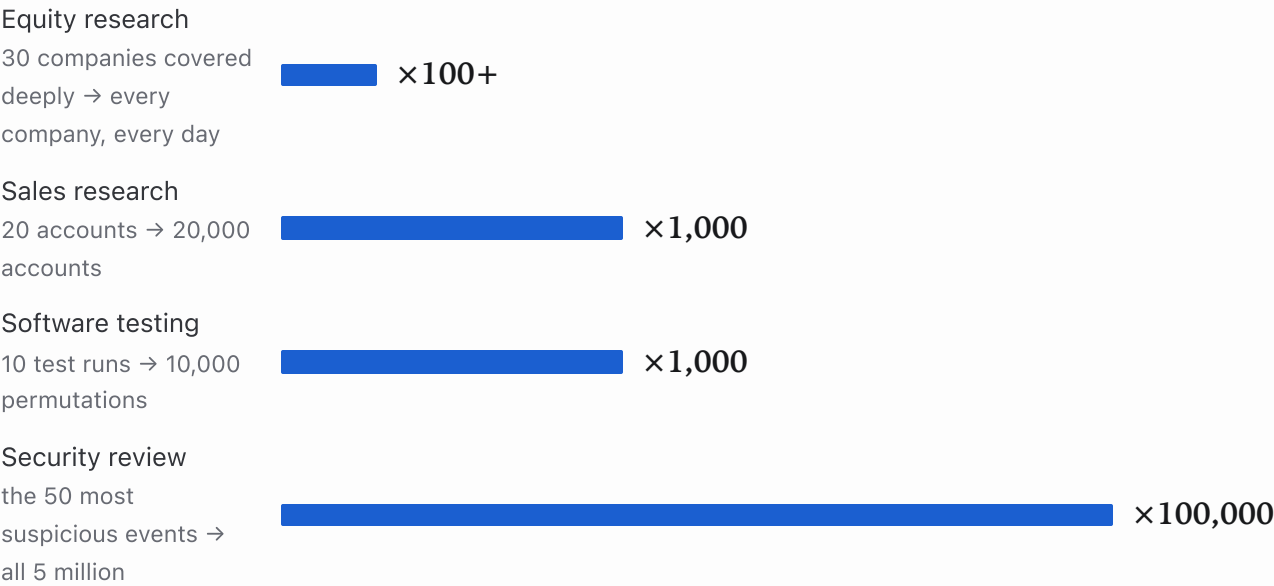
An agent can plan a step, use a tool, inspect the result and continue. The loop may finish, fail or require human help. More steps create activity; accepted results establish usefulness.

FIGURE 1.4

ILLUSTRATIVE

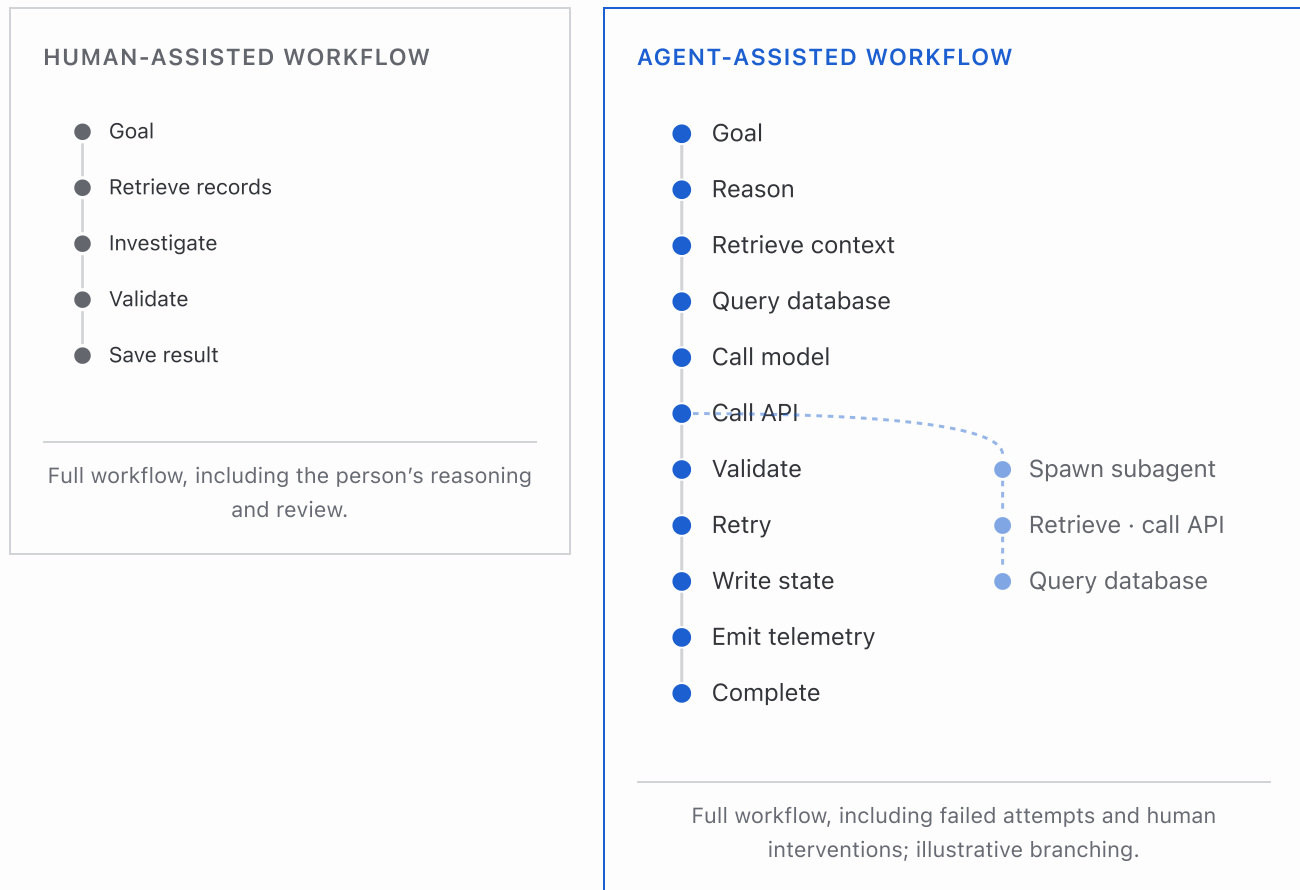
Cheaper work expands the task list

WORK PERFORMED WITH AGENTS ÷ WORK PERFORMED WITHOUT THEM, ILLUSTRATIVE, LOG SCALE



Illustrative examples show work that could become economical as the cost of using agents falls. The task counts are assumptions. They explain the mechanism behind our demand thesis rather than measure adoption in 2026.

One goal can generate many infrastructure actions



A single task can involve model calls, searches, code execution, database reads and stored results. The diagram is illustrative. Calls, tasks and machine hours are different units and should not be added together.

The evidence is strongest in software

Public GitHub activity gives us an early view. Merged code changes rose from roughly 3 million a month in 2023 to more than 18 million in September 2026. New repositories also grew sharply.

These are development events. A merged change does not establish AI authorship, quality or economic value. The series gives us a way to follow software production while better measures of agent use develop.

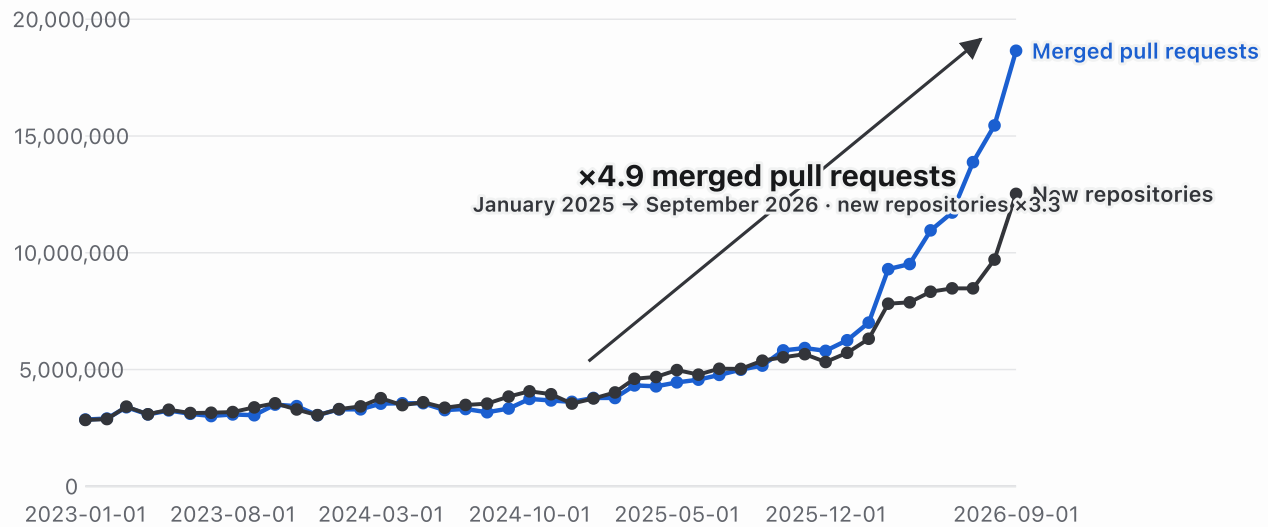
Company disclosures point in the same direction. Agents are writing code, creating databases and deploying applications. Other evidence is less encouraging: adoption remains uneven, review is costly, and some customers reduce usage after the first experiment. We keep both sides in the record.

FIGURE 1.6

MEASURED

Software production is rising

PUBLIC GITHUB · MONTHLY SEARCH COUNTS



[Follow the production series live in the tracker →](#)

Public GitHub merged pull requests rose from roughly 3 million a month in 2023 to more than 18 million in September 2026. Repository creation provides a second view of activity. These counts do not identify AI authorship or establish the quality, continued use or value of the software produced.

Early evidence cuts both ways

MACHINE LABOR IS ARRIVING

October 29, 2024

Salesforce launches Agentforce at \$2 per conversation

Machine work gets a list price · [Salesforce](#)

May 14, 2025

Databricks agrees to buy Neon; over 80% of Neon databases were created by AI agents

Agent-created state landing on a meter · [Databricks](#)

May 15, 2025

Salesforce adds a ten-cents-an-action Flex Credits option to Agentforce beside its \$2-a-conversation price

Action-based pricing arrives; the effective cost depends on actions per conversation · [Salesforce](#)

February 5, 2026

METR: the 50%-success task horizon reaches about 12 hours (718.8 minutes), the largest reading inside the suite's 16-hour reliable range

Software-task difficulty agents handle at 50% success, in human-expert hours · [METR](#)

February 6, 2026

OpenRouter: agent tokens pass human tokens on the 7-day average

Doing passes asking · [OpenRouter rankings](#)

August 2026

Merged pull requests on public GitHub pass 15M a month: about 5x in under four years, on ~25% more developers

Production bends away from its trendline · [GitHub Search API \(tracker\)](#)

September 6, 2026

OpenAI reports 3.1 agent-workdays for every human workday inside its research organization

The frontier duty cycle, observed · [OpenAI](#)

12 major events of 19 in the record, each linked to its source where one is public. The full record, including the minor rows, is on the tracker.

WHERE IT HAS FALLEN SHORT

2025

Klarna, which said its assistant did the work of 700 service agents, walks back the AI-only strategy and rehires humans for quality

Agent economics get re-priced in daylight · [Klarna](#)

June 25, 2025

Gartner predicts more than 40% of agentic AI projects will be canceled by the end of 2027 on cost, unclear value and weak controls, and names "agent washing" by vendors

A dated call against adoption, gradable at the end of 2027 · [Gartner](#)

July 10, 2025

METR randomized trial: 16 experienced open-source developers completed tasks 19% slower with early-2025 AI tools, while estimating they were 20% faster

Capability is not yet productivity for experts on mature code · [METR](#)

August 21, 2025

Commonwealth Bank of Australia reverses 45 customer-service job cuts made for an AI voice bot after call volumes rose rather than fell, and apologises

Agent economics re-priced in daylight, a second Klarna · [ABC News](#)

December 3, 2025

Salesforce reports 18,500 Agentforce deals since launch, 9,500 of them paid, in its quarter to October 31, 2025, about a year after general availability; deals are not customers, so the share of customers paying is not disclosed

Machine work is priced; paid adoption is still narrow · [Salesforce, FY26 Q3 earnings release](#)

The record combines company adoption disclosures, pricing changes and capability tests with evidence of

weak retention, incomplete deployment and costly review. Each observation carries its date and denominator. Salesforce's conversation and action prices use different units; they do not establish a 95% price cut for the same task. METR (Model Evaluation and Threat Research) measures the horizon as the human time associated with tasks an agent completes at a stated success rate, not continuous autonomous runtime.

Two paths to 2032

Our scenarios differ mainly in adoption and autonomous work. Both include agents directed by people and agents that start tasks under standing instructions.

In the **Agentic Workforce**, adoption reaches 70% of knowledge workers by 2032, a base of about 1.2 billion people in our count (Appendix A5). Most agents remain human-directed, but the smaller autonomous group contributes a large share of the hours because it runs longer. About 17 billion agents work an average of three hours a day. That produces an annual rate of about 19 trillion machine hours.

In the **Machine Economy**, adoption reaches 95%. Businesses deploy many more autonomous agents, and a larger share works through most of the day. About 68 billion agents average 13.6 active hours a day. The annual rate reaches about 340 trillion hours, subject to the scenario's compute assumptions.

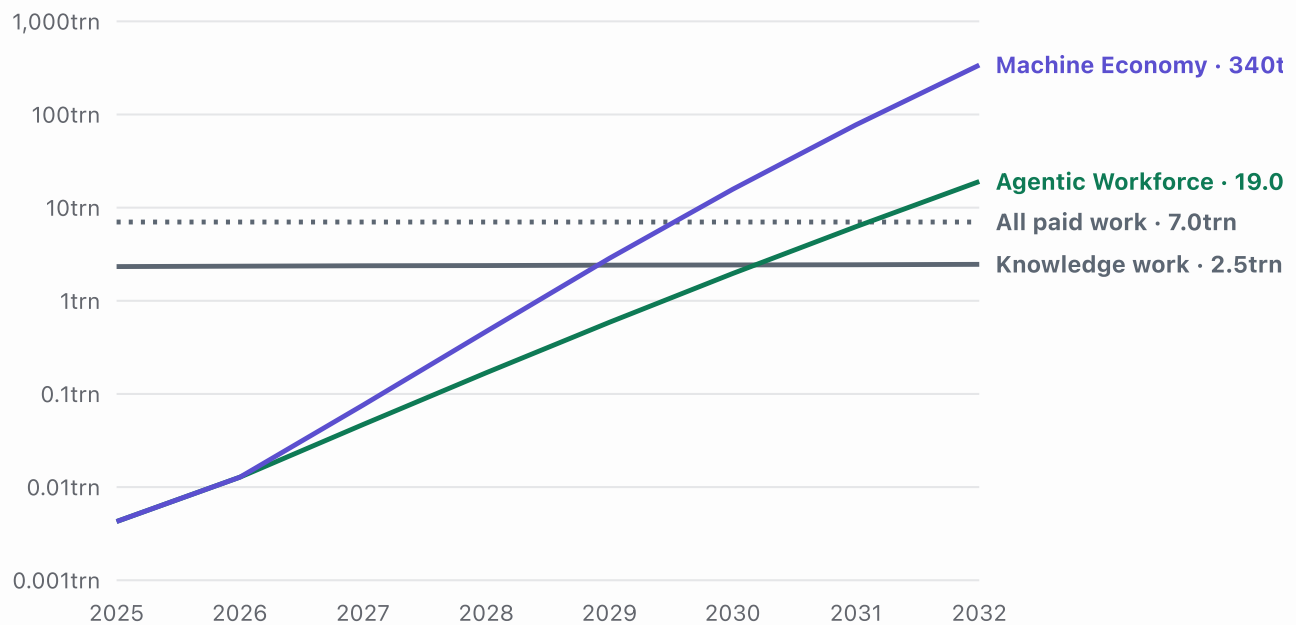
Neither is a cautious case. From the 2026 starting point, the first requires machine hours to multiply about 3.4 times a year; the second requires about 5.5 times. Slower adoption, limited repeat use or inadequate compute can produce much smaller outcomes.

FIGURE 1.8

SCENARIO

Two paths for machine hours

Machine hours a year against all paid human work and human knowledge work, log scale; move the two sliders to draw your own case



[Open the full machine hours model →](#)

The Agentic Workforce reaches about 19 trillion annual hours in 2032; the Machine Economy reaches about 340 trillion under the stated capacity assumptions. Both are modeled scenarios. The two horizontal lines are all paid human work, seven trillion hours a year, and knowledge work, about 2.5 trillion. The vertical scale is logarithmic, so equal spacing represents equal multiples. Sources and method: [machine hours model](#).

The prediction we are making

We estimate all paid human work in 2026 at about seven trillion hours a year. In our published scenarios, machine hours pass that benchmark in 2032 and 2030 respectively.

This is a comparison of time. A machine hour may contain a failed attempt, while a human hour may contain judgment an agent cannot provide. Passing the line would establish the scale of machine activity. Its usefulness and value must be measured separately.

Our central prediction is the crossing by the end of 2032. The much larger Machine Economy outcome is a stronger claim. It can fail even if the crossing occurs.

The growth required to pass human hours

What must happen for machine hours to pass all paid human work by 2032: the yearly multiplier each path carries, and the minimum combinations that clear the threshold

From 35 million machine hours a day in 2026 (0.013 trillion a year)

PATH FROM THE 2026 STARTING POINT	MULTIPLIER A YEAR, 2026 TO 2032	HOURS IN 2032	
Constrained adoption	2.42x	2.6 trn	misses paid work; reaches knowledge work in 2032
Knowledge-work threshold	2.40x	2.5 trn	the hours of 1,179 million knowledge workers
All-paid-work threshold	2.86x	7.0 trn	the headline crossover
Agentic Workforce	3.38x	19 trn	passes paid work in 2032
Machine Economy	5.46x	340 trn	passes paid work in 2030

Three minimum combinations that clear all paid human work in 2032 (people × agents × hours = 19 billion machine hours a day)

ADOPTION, SHARE OF KNOWLEDGE WORKERS	AGENTS PER USER	ACTIVE HOURS A DAY PER AGENT NEEDED
45%	10	3.6 h
70%	5	4.6 h
95%	3	5.7 h

The crossover with all paid human work needs about 2.86x a year from the anchor for six years; the Agentic Workforce carries 3.38x, so a multiplier about 15% lower than our softer scenario misses it, and the constrained case at 2.42x does. The two scenarios cross because of their assumptions, not as independent evidence that the crossing is likely; this exhibit is what the assumptions have to deliver. Scenario; the human benchmark is 7.0 trillion paid hours in 2026, held flat (A5).

Annual growth required from the original 2026 forecast starting point through 2032. These are requirements of the forecast, not observed growth rates. The benchmark is seven trillion paid human working hours in 2026.

II

Counting machine hours

How we estimate current activity and build the 2032 scenarios.

12 MIN READ · 7 FIGURES

[Counting machine hours](#) [What the public data can tell us](#) [From today's estimates to the scenarios](#)
[How many people use agents](#) [How long each agent works](#) [Capability and working time are different](#)
[Our own measurements](#) [The activity behind the OpenRouter estimate](#) [From hours to tokens](#)
[Compute can limit the outcome](#) [How we will judge the crossing](#)

Counting machine hours

One agent actively working for one hour counts as one machine hour. Parallel agents count separately; idle time does not. The model begins with a simple calculation:

Active agents × active hours per agent = machine hours.

We build the agent population from the number of people using agents and the agents associated with each user. We then separate human-directed work from autonomous work. This makes the assumptions visible, even though no representative global survey yet measures all of them.

What the public data can tell us

There is no worldwide meter of machine hours. We estimate it from disclosed model usage and from profiles of how agents work.

A token is a small piece of text a model reads or writes. Token totals include ordinary chat as well as agent work. They also include repeated context, so the same information can be processed many times during a task.

Our **provider-by-provider estimate** combines each venue's token volume, its assumed agent share and the tokens its agents process per hour. We add the hours and account for traffic routed through OpenRouter so that it is counted once.

Our **OpenRouter estimate** applies the platform's agent mix to the wider market. We separate coding, research and enterprise workflows because their token use per hour differs. OpenRouter's task classifications describe platform traffic; using them for agent applications is an assumption.

For the week ending October 18, 2026, this edition's **OpenRouter-based estimate is 55 million hours a day**; the tracker carries the current weekly reading. The alternative provider-by-provider estimate is 17 million. Neither is a direct count. They share assumptions, so their disagreement is useful information rather than independent confirmation. The original 2026 forecast starting point remains 35 million hours a day.

Changing the workload profiles gives a wide range. The OpenRouter estimate is roughly 25–154 million hours a day at the profile endpoints. That is an assumption range, not a statistical confidence interval.

From today's estimates to the scenarios

The original 2026 forecast starting point is 35 million machine hours a day, an annual rate of about 13 billion. We set the inputs on October 6, 2026 and registered the launch prediction on October 8. This starting point stays fixed under the published review rule; the tracker updates current estimates separately.

FIGURE 2.1

MODELED

Build hours from agents and active time

2026 → 2029	2026 MODEL STARTING POINT	AGENTIC WORKFORCE 2029	MACHINE ECONOMY 2029
Active agents working	142m	1.9bn	3.9bn
× Hours per agent machine hours a day	≈ 0.25 h	≈ 0.86 h	≈ 2.00 h
≈ Machine hours a day	35m	1.6bn	7.8bn
Daily machine hours × 365 days	35m	1.6bn	7.8bn
≈ Machine hours a year	12.79bn	0.59trn	2.85trn
× Revenue per machine hour four-company revenue per hour	0.29¢	0.15¢	0.06¢
≈ Agent revenue annual, four companies	\$37.0m	\$0.9bn	\$1.7bn

Active agents multiplied by daily active hours per agent gives machine hours a day; multiplied by 365 it gives the annual rate. The 2026 reading and the two 2029 scenarios show each term, and the second line prices those hours at the four company cases, which Part Three builds meter by meter. Agents are active users multiplied by agents per user, with human-directed and autonomous work calculated separately and added (Figure 2.2). Daily hours are averaged across all calendar days. Sources and method: [Appendix B1](#).

The assumptions that separate the scenarios

From people to machine hours: the build at 2032 in both scenarios, every row an assumption you can argue with

2032 ASSUMPTION / RESULT	AGENTIC WORKFORCE	MACHINE ECONOMY
AGENT USERS	1,179M workers × 70% adoption 825M users	1,179M workers × 95% adoption 1,120M users
HUMAN-DIRECTED HOURS / DAY	825M users × 16 agents × 1.8 hours 23,769M hours	1,120M users × 16 agents × 1.8 hours 32,257M hours
AUTONOMOUS HOURS / DAY	825M users × 5 agents × 6.875 hours 28,370M hours	1,120M users × 45 agents × 17.84 hours 899,176M hours
TOTAL ANNUAL HOURS	19.0 trillion	340 trillion

Annual hours = (human-directed + autonomous daily hours) × 365. M = million. Daily totals are calendar-day averages; figures are rounded after calculation. Inputs are assumptions, not observed adoption or working time.

► EVERY ROW OF THE 2032 BUILD, BOTH SCENARIOS

[Move every row on the machine hours model →](#)

The 2032 build shows adoption, agents per user and active hours for each group. The largest difference comes from autonomous work: how many agents operate under standing instructions and how long they stay busy. These inputs are scenario judgments. Registrations and software licenses do not directly measure them. Sources and method: [Appendix B1, machine hours model](#).

How many people use agents

The model assumes that 12% of knowledge workers actively use agents in 2026. By 2032, that share reaches 70% in the Agentic Workforce and 95% in the Machine Economy.

Surveys help us judge the direction, but they answer different questions. Census measures AI use among US businesses. Gallup measures use among US employees. Ramp observes payments among its customers. A paid license, a chatbot visit and an agent used repeatedly in production are different observations.

We use these sources as checks on our assumptions. None establishes the worldwide share of knowledge workers using agents. The faster scenario assumes deployment spreads through firms as well as through individual users.

How long each agent works

This is the largest uncertainty in the hours forecast. Both scenarios reach sixteen human-directed agents per user, each working 1.8 hours a day in 2032. That is 28.8 machine hours per user per calendar day, spread across parallel agents.

The autonomous assumptions differ much more. The Agentic Workforce reaches five autonomous agents per user. A quarter of them work 17 hours a day; the rest average 3.5 hours. The Machine Economy reaches 45 autonomous agents per user. Of these, 70% work 23 hours a day; the rest average 5.8 hours.

These are strong assumptions about useful, affordable activity. An agent with a permanent identity is not necessarily busy. A business will keep running it only if its work is worth the cost and its mistakes can be controlled.

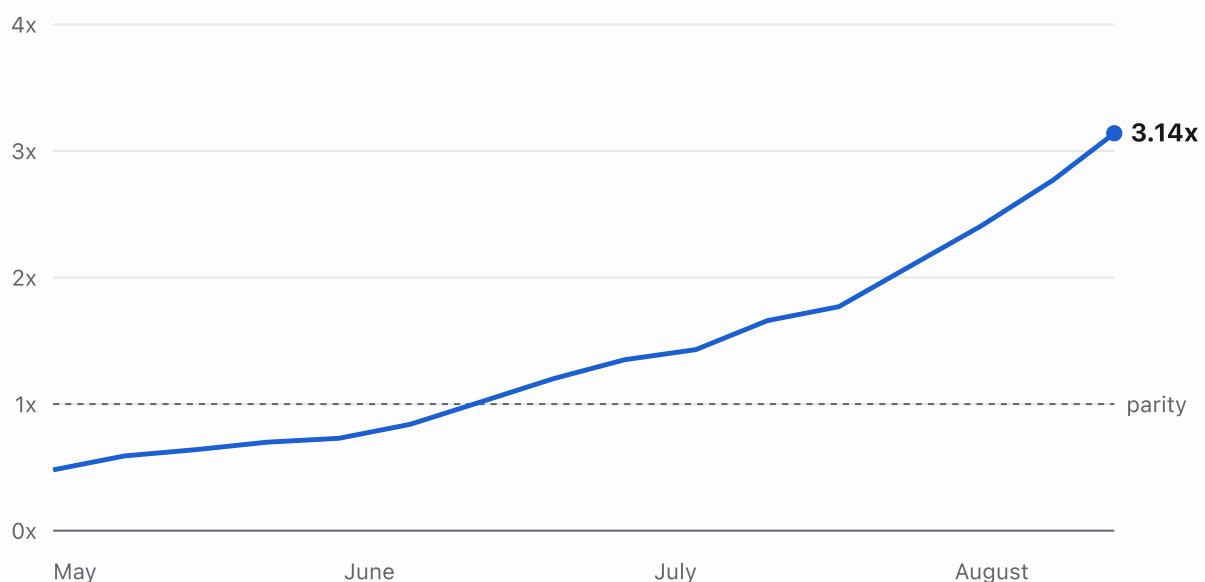
OpenAI's account of its research organization provides one comparison. It reported 3.1 agent workdays per human workday in mid-August 2026. That shows what an intensive user can do; it does not establish what an average business will do. Its definition of an agent workday also differs from our definition of an active hour.

FIGURE 2.3

OBSERVED · REPORTED, TRACED

Agent workdays per researcher workday at OpenAI

Agent workdays divided by researcher workdays inside OpenAI's research organization, May to mid-August 2026, traced from the company's chart



OpenAI reported 3.1 agent-workdays per human researcher workday in mid-August 2026. The chart approximately redraws the company's May-to-August series. The 3.1 figure is a ratio of agent days to human days, not a growth multiple. The company's definition may include time excluded from our active-hour measure. Source: [OpenAI, Research acceleration: the view inside OpenAI](#).

Capability and working time are different

METR measures the difficulty of tasks an agent can complete at a stated success rate. Difficulty is expressed as the time a human expert would need. An eight-hour task on that measure might take an agent twenty minutes.

We therefore track three quantities separately: time agents spend working, tasks they complete, and the difficulty of those tasks. More capable models can attempt harder work and complete familiar work faster. The net effect on hours depends on how much additional work users start.

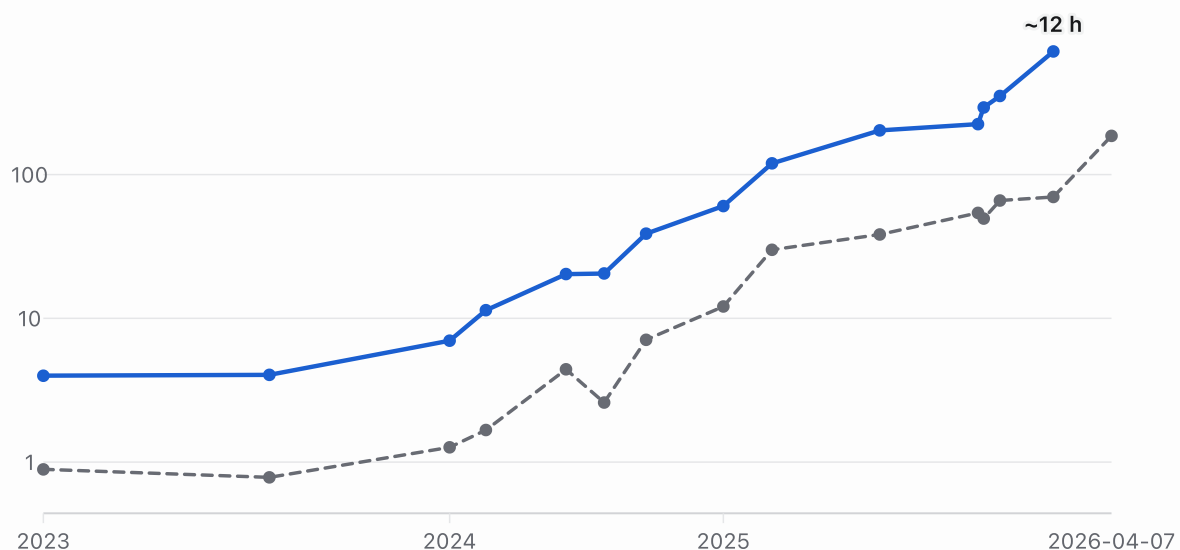
FIGURE 2.4

MEASURED

Agents are completing longer benchmark tasks

HUMAN-EXPERT TASK MINUTES · LOG SCALE

- Task length finished half the time (50% success)
- - - Task length finished four times in five (80% success)



[Follow the capability series live in the tracker →](#)

METR estimates the human-expert time associated with tasks that a model completes at 50% or 80% reliability. Higher values indicate success on longer tasks in that evaluation. They do not establish equal reliability in business workflows or the duration an agent ran. Read the success rate, evaluation date and uncertainty together. This is the historical Time Horizon 1.1 series. METR marked its source page as no longer actively updated on September 8, 2026; a recent fetch is not a new evaluation. Central estimates above 16 human-expert hours are omitted as unreliable under this task suite, so a line ends where its next reading would exceed that limit. Sources and method: [benchmark_results_1_1.yaml](#).

Our own measurements

We use our agents' logs to study what an hour contains. The coding logs show heavy reuse of cached context. Our separate workflow experiment ran a service agent through 48 benchmark tasks. It passed 52% of them and processed about 13 million tokens per active hour.

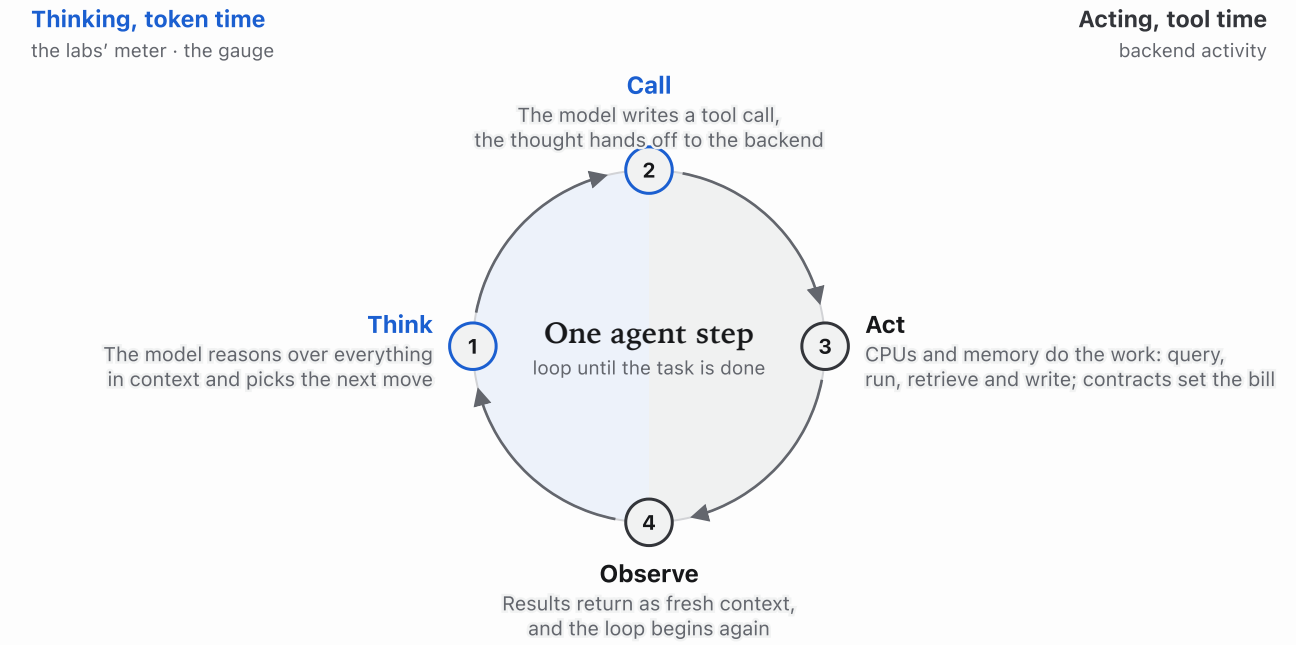
The workflow trace added about 1.07 million fresh input tokens and wrote about 0.32 million output tokens per hour. These are separate quantities. The experiment is small, and the customer was simulated. It calibrates one profile; it does not measure enterprise productivity as a whole.

Since October 7, 2026, our published coding series counts the time a model generates or a tool runs. It excludes waiting for a person and counts a parent's wait for a child only in the child's work. The earlier estimate, based on capped gaps between calls, remains in the record. The change produces fewer measured hours and more tokens per hour; it is a change in method.

The workflow panel brings hours, tokens, tool use and review measures together from the same runs. It includes our coding sessions, the service experiment and a web-searching agent tested on short factual questions. A public OpenHands evaluation adds a second coding operator. Each pass rate belongs to its own benchmark; none is applied to global hours.

Vendor disclosures help us follow agent-built applications from creation to deployment and paid use. They do not yet show how many still generate bills after ninety days. That remains a gap in the evidence.

From model use to infrastructure use



- 1

Workload

A task creates execution, state and telemetry.
- 2

Offsets

Caching, batching, sampling, local execution and substitution.
- 3

Contract

Free tiers, commitments, resource units and realized prices.
- 4

What reaches a meter

Which of it a vendor can bill is Part Three.

Combined exhibit: the panels above are shown together as one figure.

Thinking, tool use and execution can each draw on different services. Caching, batching, local execution and bundled allowances can reduce the paid bill. The diagram explains the paths through which activity may become revenue; it does not assign a fixed infrastructure cost to every hour.

The activity behind the OpenRouter estimate

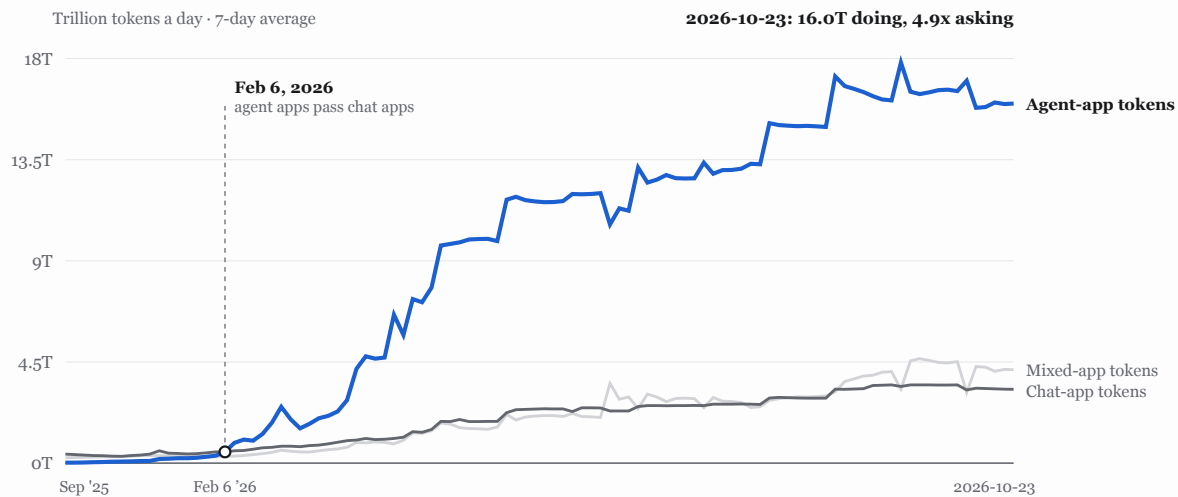
OpenRouter's application mix helps us judge how much of its traffic comes from agent work. Applying that mix to the wider market remains an assumption. The tracker shows the latest reading alongside the provider-by-provider estimate.

FIGURE 2.6

OBSERVED · CLASSIFIED MIX

Agent applications gain share on OpenRouter

Tokens a day on OpenRouter, 7-day average: agent-app, chat-app and mixed-app tokens



[Follow this live on the tracker →](#)

The historical trace shows agent-app token use passing chat-app use around February 6, 2026. Earlier points were read from the published chart through August 7, 2026; later observations use application classification; the two series join at August 7, 2026. This measures traffic on OpenRouter, whose mix may differ from the wider market.

From hours to tokens

The model uses about 1.4 million fresh tokens per machine hour in 2026, blended across coding, research and enterprise workflows. This rate updates with the data and is recorded month by month. The original 2.3-million figure describes the coding sample used to set the hours anchor; it is kept in the record.

At roughly 800,000 fresh tokens per hour in 2029, the Machine Economy's 2.9 trillion hours imply about 2,300 quadrillion tokens a year. The fall from about 1.4 million fresh tokens an hour today assumes that routine steps move to smaller models and that more context is served from cache as

tools mature; it is an assumption, not a measured trend. That is about 130x the 2026 base of about 17 quadrillion. The calculation translates the hours scenario into token demand. It is not a separate forecast of revenue.

Compute can limit the outcome

Demand does not guarantee that enough capacity will exist to serve it. We compare the hours forecast with an illustrative compute build, allowing for differences in fresh input, cached context and written output.

On the published assumptions, the 2029 build can serve about 4 trillion machine hours a year. A price-weighted stress test puts it nearer 3 trillion. Token prices are not physical compute ratios, so neither calculation is a measured global ceiling.

The Machine Economy's 340 trillion hours use about **95% of the illustrative 2032 capacity allowance** of 357 trillion hours. It requires compute to double each year after 2029, tokens per hour to fall tenfold, and nearly the full allowance to be available to agents. The stricter case also needs less repeated context for each fresh token. If those changes fall short, delivered hours fall below demand. These assumptions describe a demanding supply path; they do not establish its physical feasibility.

We also check the scale against estimates of installed AI chips and disclosed inference throughput. That hardware comparison in Appendix A3 is a plausibility check. It depends on how much capacity serves inference, how heavily it is used and which models it runs.

The model shows requested work, available capacity and delivered work separately. Revenue follows delivered hours. The share of capacity available to agents is adjustable because agents compete with other uses of AI.

How we will judge the crossing

The headline prediction uses the last complete week of December 2032. We multiply average daily machine hours by 365 and compare that annual rate with seven trillion paid human hours.

A robust success requires both estimates—the provider-by-provider build and the OpenRouter extrapolation—to clear the top of the human-work benchmark range: even the lower end of each sensitivity range must exceed 7.5 trillion hours a year. It is not enough for their central estimates to pass seven trillion. Appendix D2 sets the full rules for a success, a miss or an unresolved result. Later activity cannot rescue the prediction; better evidence can revise the estimate of the original observation week.

The model's growth paths will change as evidence improves. The registered prediction remains on the record.

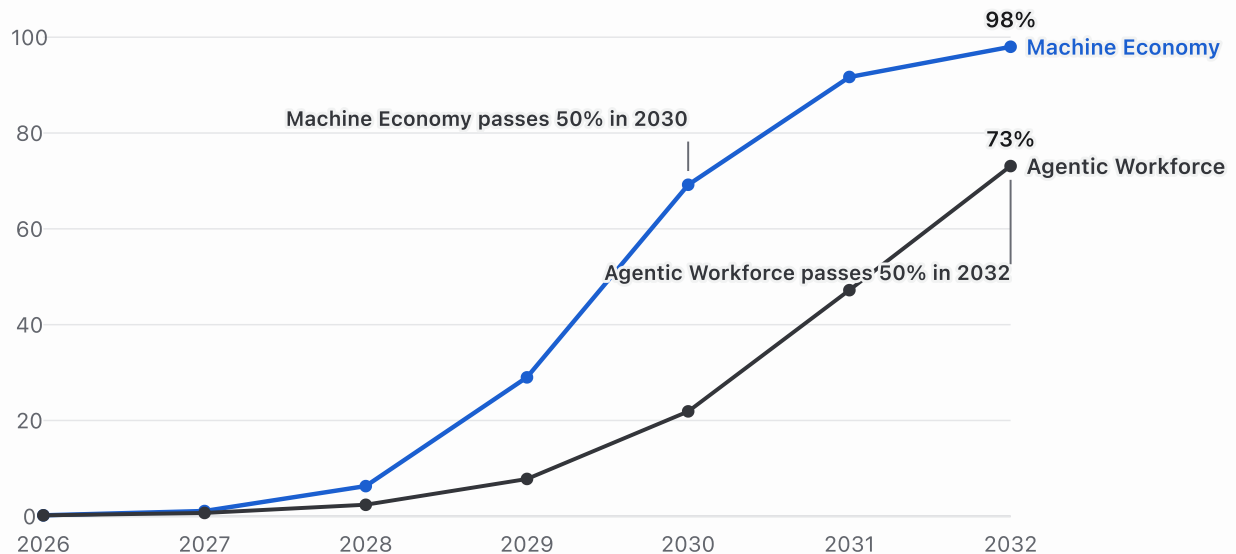
FIGURE 2.7

SCENARIO · GRADED QUARTERLY

Machine hours as a share of all hours worked

The machine share of all hours worked, both scenarios; move the sliders to add your own case

MACHINE SHARE OF ALL HOURS WORKED, %



[Open the machine hours model behind this figure →](#)

Divide annualized machine hours by machine hours plus seven trillion paid human hours. A value above 50% means machine activity exceeds the human benchmark in duration; the Machine Economy passes 50% in 2030 and the Agentic Workforce in 2032. It does not imply equivalent output, productivity or jobs displaced. The benchmark is held constant throughout the comparison.

III

The infrastructure behind the work

How agent activity reaches infrastructure and when that use becomes revenue.

12 MIN READ · 10 FIGURES

[The infrastructure behind the work](#) [Five layers of the system](#) [Who serves the new demand](#)
[Four ways an hour reaches a bill](#) [What the companies have reported](#) [From working time to revenue](#)
[The opportunity beyond current software budgets](#)

Research commentary from [Skycatcher](#), LLC, an exempt reporting adviser; not an offer, solicitation, recommendation or investment advice. The receipts, list prices and revenue paths in this part are our illustrations of the scenarios on stated assumptions, not company guidance and not price targets, and the companies named have not reviewed them. [Skycatcher](#), its affiliates, principals and clients may hold long or short positions in any public company named here and may trade them without notice. [Appendix D3](#) carries the full disclosures.

The infrastructure behind the work

Agents create demand while they work—and through the software they leave running. A task can create a database, an application or a recurring process that remains in use after the agent stops.

Lasting revenue depends on continued use and payment. We follow three steps: more agent activity, more infrastructure used, and more revenue earned by providers. Each step needs its own evidence.

The opportunity spans cloud platforms, established providers, open-source systems and new companies. Customers can choose bundled services or build their own. The four company cases below show particular ways activity can become revenue.

The sizing figure at the end of this part compares roughly \$190 billion in selected backend categories with the \$25 trillion to \$35 trillion paid for knowledge work each year. The categories have different scopes and may overlap; the wage comparison establishes scale, not an addressable market.

Five layers of the system

Computation supplies chips, servers, networks and energy. **Intelligence** supplies the models. **State** stores data and memory. **Work** runs and coordinates software, controls access and monitors activity. **Applications** deliver the result to the user.

We call State and Work the backend. The labels help organize the research; products often cross the boundaries.

A model can propose an action. The systems around it must decide whether the action is allowed, apply it consistently and preserve a record. Better databases do not make an agent's reasoning correct. They help contain mistakes, inspect what happened and recover.

The services behind agent work

PUBLIC PURE PLAYS PER LAYER · THE UNIVERSE IS L2 AND L3

L4	Applications Where the outcome lands and the seat is billed; business software only, consumer and fintech out.	49 MAPPED · OUTSIDE THE UNIVERSE
L3	Work Executes, coordinates and accounts for machine activity: execution and delivery, developer platforms, observability, identity, security of the runtime, data protection.	27 BACKEND · THE UNIVERSE
L2	State Stores records, memory and controlled business data. Integration and existing data can make a service harder to replace.	5 BACKEND · THE UNIVERSE
L1	Intelligence Tokens and inference, priced, competitive, deflating; the labs are private or inside the hyperscalers, so the two listed pure plays are in Hong Kong.	2 MAPPED · OUTSIDE THE UNIVERSE
L0	Computation Silicon, memory, networking and optics, systems, power and cooling, rented capacity; the AI build is the majority of revenue and the growth is already reported.	25 MAPPED · OUTSIDE THE UNIVERSE
L0 to L3	Hyperscalers and SpaceX Amazon, Microsoft, Alphabet, Oracle, Alibaba, Baidu and Tencent sell every layer bundled and are measured on their cloud segments only. SpaceX spans Intelligence, Applications and Computation through xAI, Cursor and its data centers and reports no AI segment yet. The seven hyperscalers are drawn as a context line in Figure 4.1; SpaceX, listed on June 12, 2026, reports no AI segment and is in no index. None is in the universe or its medians.	8 CROSS-LAYER REFERENCE GROUP

Our research map follows five layers: State, Work, Applications, Intelligence and Computation. Companies can operate across several layers. The categories describe what is sold; they do not establish which businesses will capture the most value. Sources and method: [Appendix C1](#).

Who serves the new demand

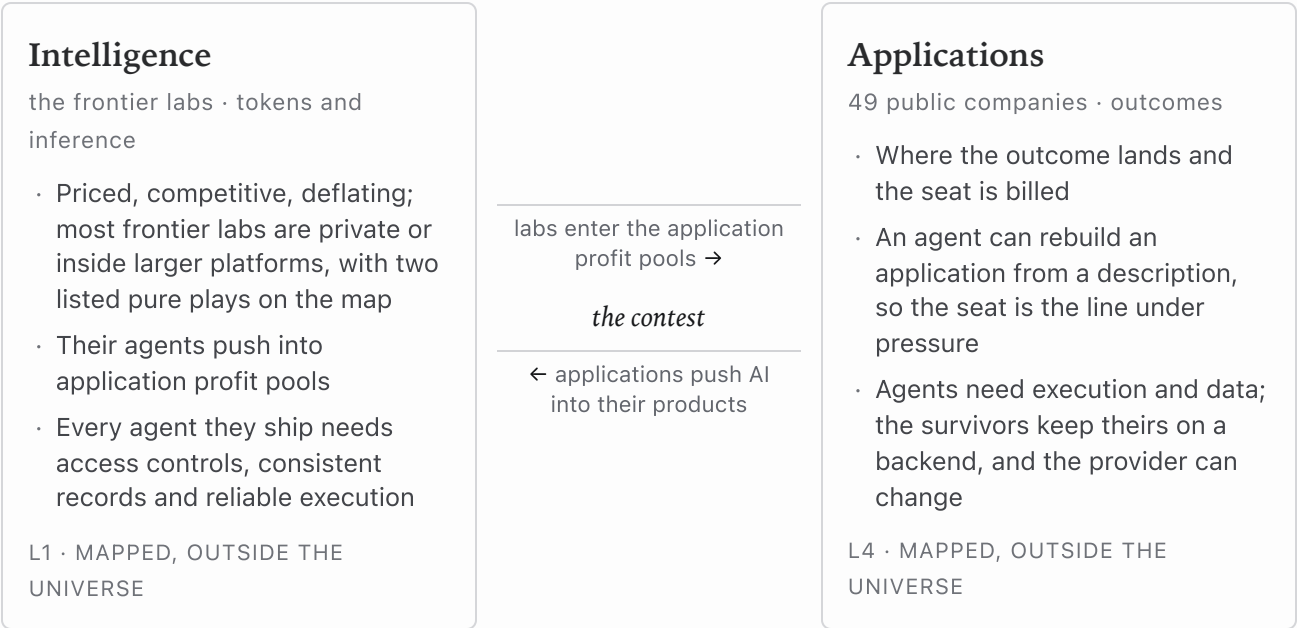
Our public research universe contains thirty-two backend companies. The wider map includes hardware, models and applications for context. The universe is a research scope, not a portfolio or a complete list of potential winners.

Cloudflare, MongoDB, Snowflake and Datadog are four detailed case studies. They let us examine different ways agent activity can become paid usage. The largest cloud platforms, private companies and open-source alternatives remain part of the opportunity even where disclosure makes them harder to model.

The distinction between State and Work is economic as well as technical. Stored data and deployed applications can persist after an agent finishes. Execution and monitoring tend to vary with activity. Both can generate recurring spending, but neither guarantees it.

Where competition can change the bill

Models and applications compete to serve users; both depend on infrastructure



↓ Agent work needs execution and data; its use of these providers varies by task ↓

The backend · the universe

32 public companies · State (5) keeps the record · Work (27) executes, coordinates and accounts for the hour · serves both sides and competes with bundled and internal systems

4 BELLWETHERS · THEIR OWN READINGS, FROM THE INVESTOR DAYS IN APPENDIX E

Cloudflare AI agent requests up more than 1,700% in the year to May 31, 2026. Requests are the meter in most contracts, and the labs' agents and the applications' agents pay the same one.

MongoDB The managed MCP server grew 8x in seven weeks from Claude Code, Codex and Cursor. Agent workloads are bursty and intense, so capacity is bought for the peaks.

Datadog Spans sent to AI Observability up more than 10x in six months, on the path from monitoring to remediation with no person in the loop.

Snowflake In Snowflake's reported test, adding runtime context improved answer accuracy: 86% on hard questions, compared with 24% for a frontier coding agent using the raw SQL interface. This is a company-reported result on its test.

L2 + L3 · 32 PUBLIC COMPANIES · MONGODB AND SNOWFLAKE BILL STATE, CLOUDFLARE AND DATADOG BILL WORK

Model providers, applications, cloud platforms and independent infrastructure companies can each serve part of an agent's work. Integration may shift spending between them. The map shows possible relationships, not fixed winners or market shares. Sources and method: [Appendix E](#).

State remembers and Work executes

State

Memory and operational records

- Potential advantage: governed data and integration
- Measure: net retained bytes, resource consumption and customer expansion
- Risk: portability, compression and alternatives

Work

Execution and accountability

- Potential advantage: embedded workflows and operating reliability
- Measure: paid execution, host and telemetry resources
- Risk: sampling, bundling and runtime substitution

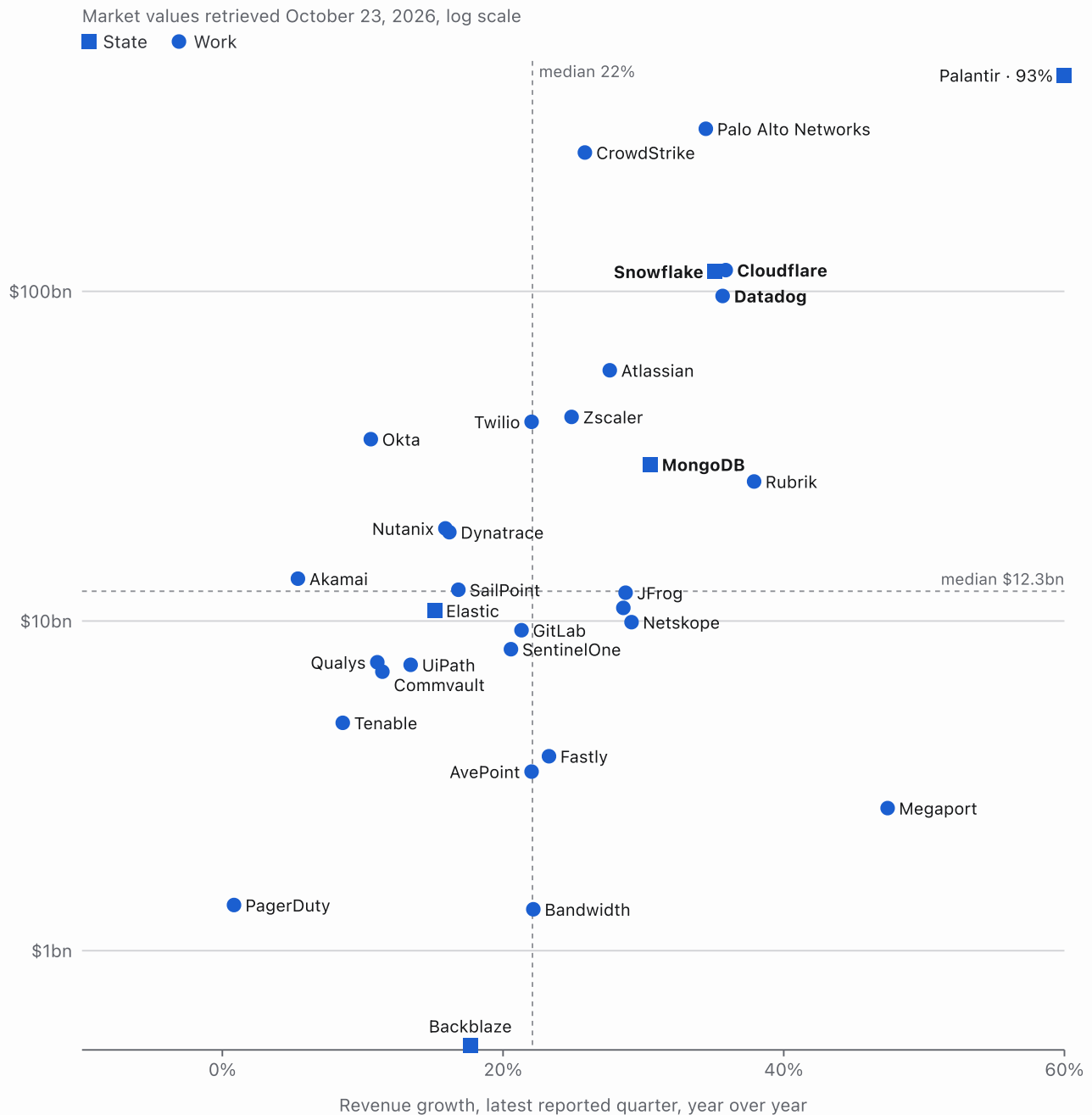
State services store the information an agent needs and the results it creates. Work services run tasks, move data and monitor execution. Both may benefit from more activity, but the financial outcome depends on pricing, utilization and whether the customer pays for the additional use.

FIGURE 3.4

OBSERVED

The public backend universe

The 32-company backend by revenue growth and market value



Revenue growth is the latest reported period against the same period a year earlier, from SEC filings; market values are the latest available observations retrieved from Financial Modeling Prep on October 23, 2026. The latest quote date recorded is October 23, 2026; individual observations can differ. Retrieval dates are not closing-price timestamps. State is drawn as squares and Work as circles. The dashed lines are the medians of the 32 companies drawn. Median growth is 30% for State and 22% for Work. Megaport and SailPoint report half-yearly, so the growth shown is the half-year to the latest period end. Palantir grows 93% and is pinned at the right edge. Unlabelled where names would overlap: DigitalOcean.

Thirty-two companies provide a broad view of the category. Revenue observations retain their reporting periods; market values retain their quote dates. The roster is a research sample, with businesses of different sizes and exposure to agents. It is not a portfolio or a measure of the entire backend market.

Four ways an hour reaches a bill

Cloudflare serves network requests and runs code. Additional activity can fit inside an existing allowance before it creates an additional charge.

MongoDB supplies database capacity. An agent’s extra reads and writes may cost the customer little until they require a larger cluster or more resources.

Snowflake charges for warehouse compute and related services. Cost depends on the queries, the time resources stay active and the use of AI services.

Datadog charges for infrastructure monitoring and selected volumes of logs, traces and other records. Customers can change how much they collect, retain and index.

These contracts explain why activity and revenue can move at different times. More calls do not become more revenue one for one. Spare capacity, caching, volume discounts and alternative products all affect the bill.

FIGURE 3.5

STRUCTURE

Four businesses and the units they bill

	CLOUDFLARE	MONGODB	SNOWFLAKE	DATADOG
Activity of the hour	Executes the work: requests, tool calls, sessions	Reads, writes and keeps state	Asks questions of governed data	Leaves traces, logs and model calls
What the meter counts	Requests against a monthly cap; CPU time; Durable Object sessions	Cluster hours by tier; Flex \$8 to \$30 a month	Warehouse seconds; AI Credits per token since April 2026	Hosts at the monthly high-water mark; bytes ingested; events indexed; LLM spans

	CLOUDFLARE	MONGODB	SNOWFLAKE	DATADOG
List price	\$2.30 per million requests all-in; \$1.44 per thousand agent sessions	M30 54¢ an hour; Voyage 2¢ to 12¢ per million tokens	XS warehouse 1 credit an hour, \$3 Enterprise; AI Credit \$2.00	\$15 to \$23 a host; \$0.10 a GB; \$1.70 per million events; \$3.50 per 10,000 LLM spans
One agent action	0.0002¢ a request; 0.14¢ for the session it runs in	\$0 until the cluster fills, then a tier step	0.2¢ busy, 5¢ if it wakes the warehouse; 49¢ per CoWork prompt	0.04¢ per traced tool call, 93% of it the LLM span
Disclosed volume	30 trillion transactions a day (June 2026)	3 trillion queries a day (June 2026)	6.3 billion queries a day (October 2024)	100 trillion events a day (October 2025)
Revenue per unit	\$0.20 per million transactions	\$2.07 per million queries, down about 30% since 2023	0.16¢ a query	\$0.12 per million events

Cloudflare, MongoDB, Snowflake and Datadog connect activity to revenue through different contracts and meters. The examples use public list prices checked on September 29, 2026. Customer discounts, commitments, included use and product mix can make realized bills materially different.

What the companies have reported

At the October 7, 2026 reading, all four case studies grew revenue faster in their latest reported quarter than in the quarter before: approximately 36% at Cloudflare and Datadog, 35% at Snowflake and 30% at MongoDB.

Their AI disclosures are less comparable. Snowflake disclosed a \$100 million AI revenue run rate in December 2025 and later discussed AI's contribution to growth. Datadog has reported the contribution of customers whose businesses center on AI. MongoDB has described growing agent activity without providing a separate AI revenue figure. Cloudflare has reported rapid developer-platform and agent-traffic growth.

These are signs of demand, with different definitions. Revenue from an AI company is broader than revenue caused by agents. A count of customers using AI is broader still. We keep those distinctions in the tracker.

The counterevidence matters. On August 6, 2026, Datadog reported lower usage from its largest customer and included that reduction in its third-quarter and full-year guidance. The customers that adopt first can also be the first to build alternatives or negotiate lower prices.

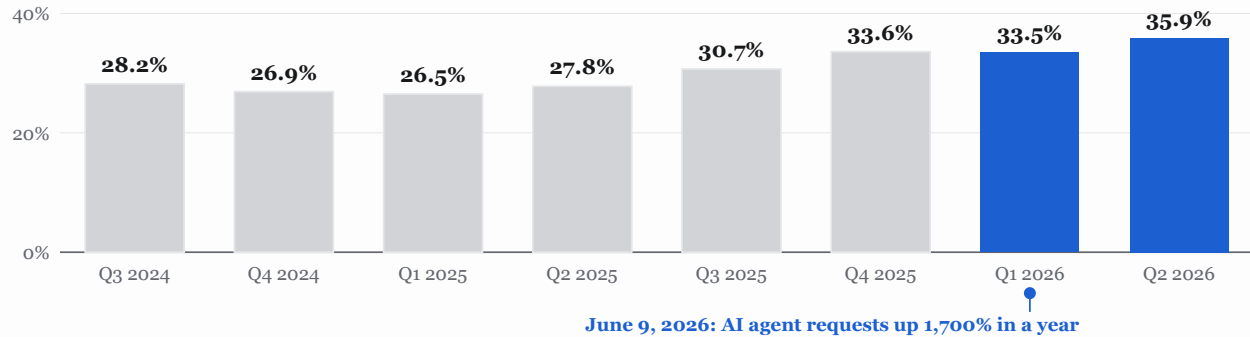
FIGURE 3.6

OBSERVED · REPORTED

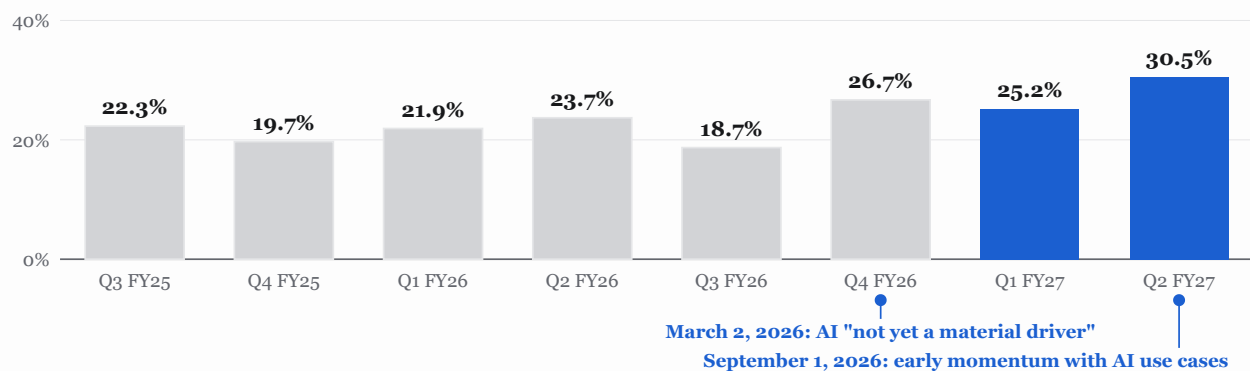
What the companies have reported

Total revenue growth, year over year, last eight fiscal quarters from SEC filings; the last two quarters in blue; the dated lines are management's AI statements on the calls

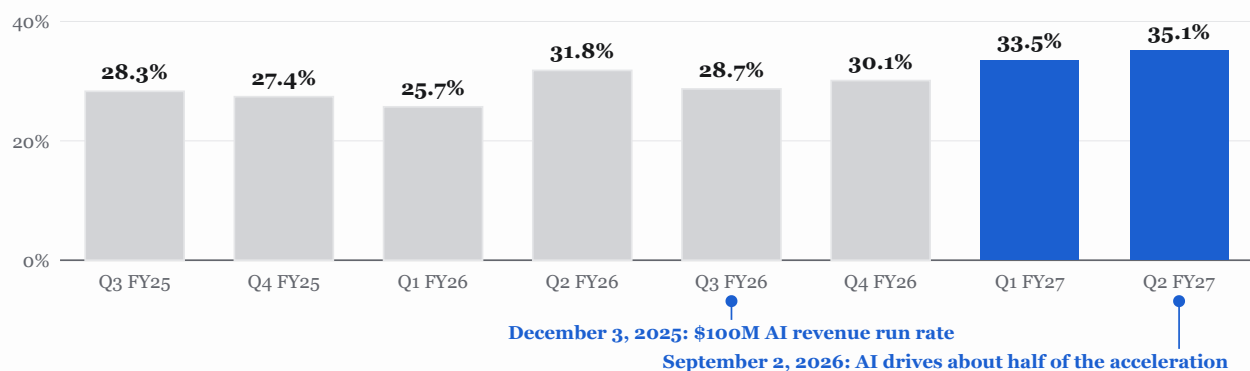
Cloudflare



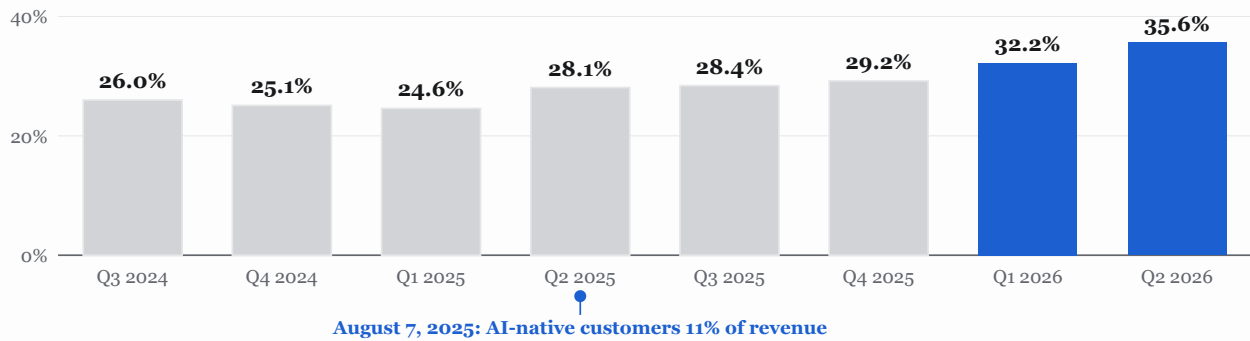
MongoDB



Snowflake



Datadog



[Follow the bellwethers' revenue growth live on the tracker →](#)

Eight quarters of revenue sit alongside dated AI and agent disclosures. The financial series establishes company performance; the disclosures provide context. Their timing alone does not show how much growth came from agents.

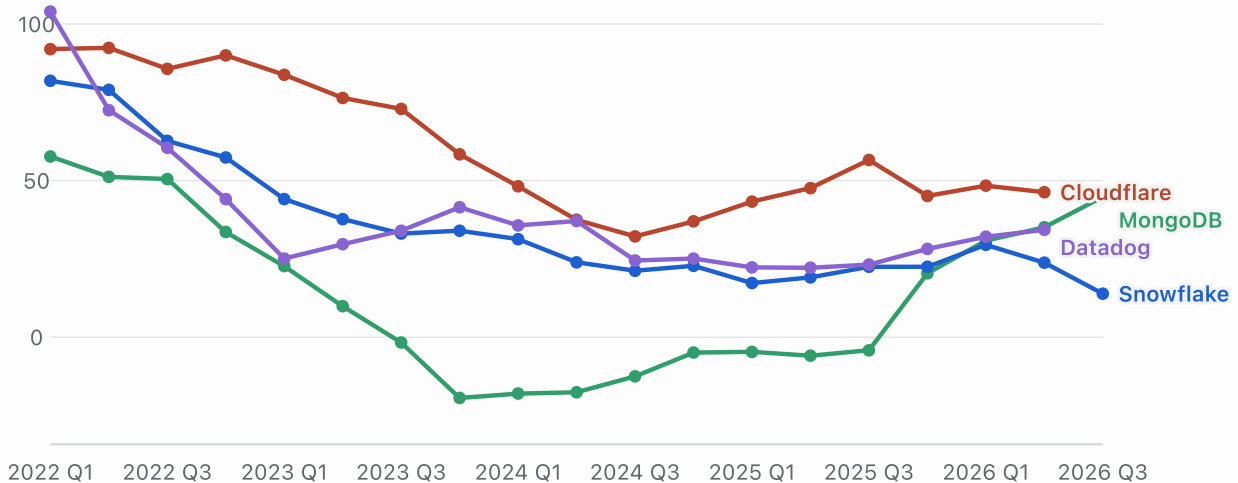
FIGURE 3.7

OBSERVED

Commitments can appear before revenue

What customers have committed to and paid ahead of use, at Cloudflare, Snowflake, MongoDB and Datadog, from SEC filings

DEFERRED REVENUE GROWTH, YEAR OVER YEAR, %



[Follow the bellwethers' deferred revenue live on the tracker →](#)

Deferred revenue helps track customer commitments and revenue timing; the tracker shows deferred revenue alongside reported revenue growth. The series covers the broader business and is affected by contract length and billing terms. It is not a direct measurement of agent demand.

From working time to revenue

For each company, we estimate:

Delivered machine hours × share using the provider × revenue per hour served.

The share using a provider is an assumption. Most global agent hours do not reach any particular company in our four-company model. One hour can also use several providers, so their shares are not exclusive market shares.

Revenue per hour served depends on the workload and contract. Our illustrative bills range from fractions of a cent for some network or database uses to between \$5 and \$33 for a data-intensive warehouse workflow. They use published prices and stated activity assumptions, before negotiated discounts.

We then let the price per comparable unit fall as volume grows. In the published scenarios, a tenfold increase in volume leaves 50% to 65% of the earlier unit price, depending on company and scenario. These are assumptions to test, not measured laws.

How one workflow could reach four bills

AN EDGE AGENT’S HOUR AT CLOUDFLARE · FETCHES, CALLS TOOLS AND RUNS ITS OWN CODE AT THE EDGE

Capacity The Durable Object session the agent runs in, and the CPU it uses: \$1.44 per thousand agent sessions on the published TCO about 0.1¢ an hour	+	Events 300 requests through the edge and Workers, \$2.30 a million all-in; most contracts cap requests a month and step up as usage runs ahead 0.07¢	+	Tokens AI Gateway and Workers AI only when the model call routes through Cloudflare; most do not 0 to 1¢	=	Illustrative hourly bill A tenth of a cent to five cents per modeled hour 0.1¢ to 5¢
---	---	---	---	---	---	---

A CODING AGENT’S HOUR AT DATADOG · RUNS ITS TESTS ON A MONITORED HOST

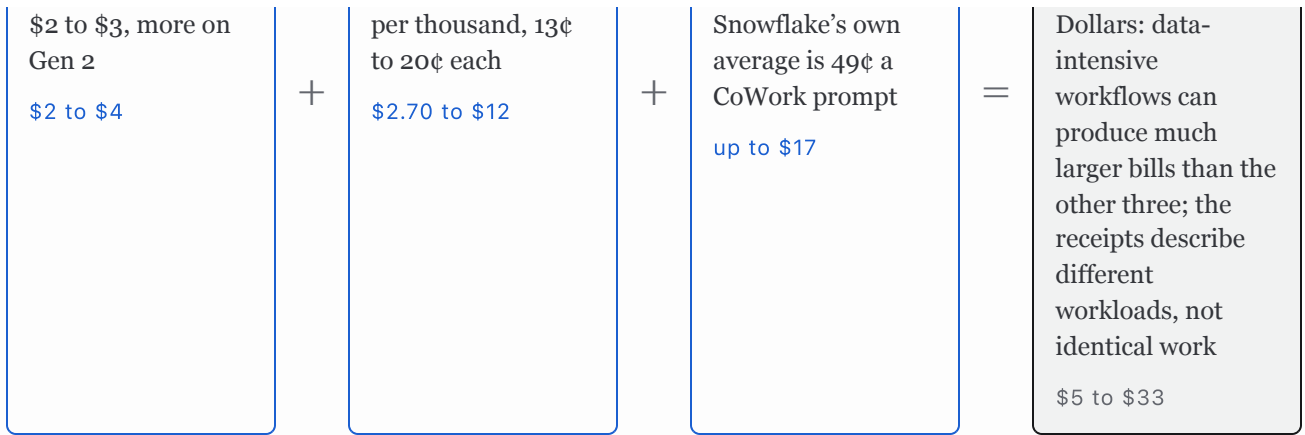
Capacity The host the agent runs on: \$15 infrastructure + \$31 tracing a month, counted at the monthly high-water mark about 6¢ an hour	+	Events 30,000 test-output lines and spans, ingested at \$0.10 a GB and indexed at \$1.70 a million 0.1¢ to 5¢	+	Tokens Every model call traced as an LLM span at \$3.50 per 10,000 after the first 40,000 a month 0 to 28¢ · 800 calls	=	Illustrative hourly bill Cents, and the model calls are most of it when they are traced 6¢ to 39¢
---	---	--	---	---	---	--

AN APPLICATION AGENT’S HOUR AT MONGODB · A THOUSAND STEPS OF READS, WRITES AND MEMORY

Capacity Share of a cluster already paid for by the hour; one more read or write is free until the tier fills 0.1¢ to 24¢	+	Events No separate per-operation charge in this example; that tier was retired on January 22, 2026 No separate charge	+	Tokens Query embeddings and reranks through Voyage, 2¢ to 12¢ per million tokens, first 200 million free 0 to 6¢	=	Illustrative hourly bill A tenth of a cent to thirty cents, depending on the workload 0.1¢ to 30¢
--	---	--	---	---	---	--

A DATA AGENT’S HOUR AT SNOWFLAKE · TWENTY TO SIXTY QUESTIONS OF GOVERNED DATA

Capacity The warehouse kept awake for the hour: one credit at		Events Cortex Analyst messages at 67 platform credits		Tokens AI Credits for the orchestration and the answer;		Illustrative hourly bill
---	--	---	--	---	--	---------------------------------



[Open the revenue model behind this figure →](#)

Illustrative receipts show where a workload might create charges at the four companies. They use stated workload and price assumptions. They are not a production customer's invoice, and they should not be summed into a universal cost per machine hour.

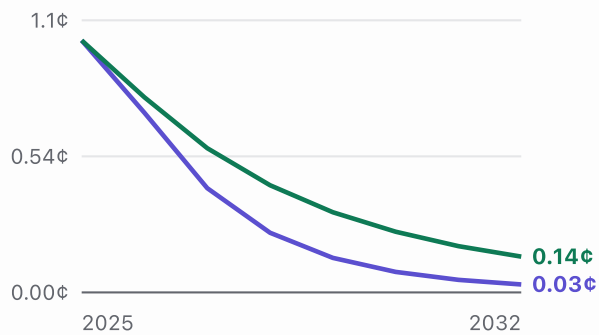
FIGURE 3.9

SCENARIO

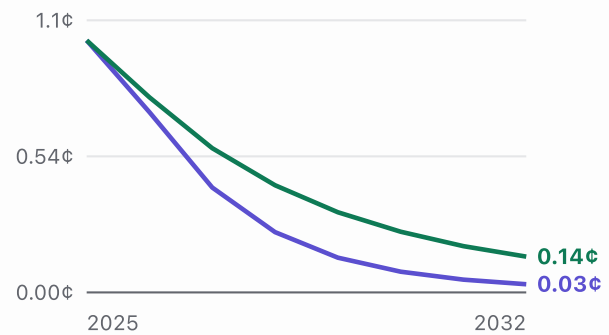
More volume can come with a lower unit price

Revenue per hour using each company, 2025 to 2032, as unit prices fall with volume

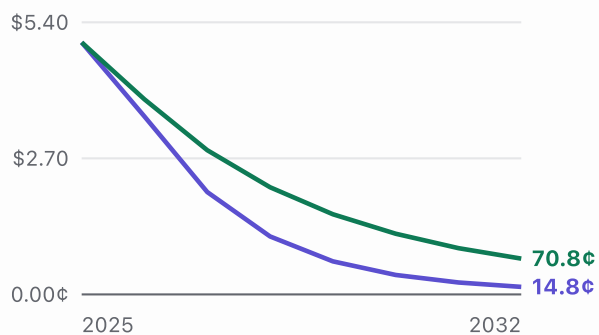
Cloudflare · 1.0¢ an hour today



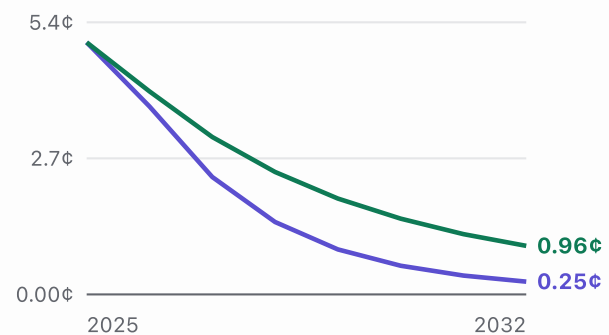
MongoDB · 1.0¢ an hour today



Snowflake · \$5.00 an hour today



Datadog · 5.0¢ an hour today



Machine Economy · Agentic Workforce. Each tenfold increase in hours using a provider reduces its revenue per hour to 50%–65% of the previous level, depending on the scenario and provider. The Machine Economy assumes more hours at a lower price per hour.

[Move the price inputs on the revenue model →](#)

Illustrative revenue per hour using each provider, 2025–2032. Prices retain 50% to 65% of their earlier level after each tenfold increase in relevant volume; this is not an annual price decline. Method and assumptions: [revenue model](#).

The opportunity beyond current software budgets

We use roughly \$190 billion as an illustrative grouping of backend markets. Their scopes differ and may overlap, so this is not a precise total for our 32-company universe.

Databases and analytics make up the largest part: about \$161 billion in Gartner's 2026 database-management forecast, compared with a \$119.7 billion market in 2024. The 2026 figure is a forecast, not reported revenue (Forecast: Database Management Systems, Worldwide, 2023 to 2029, 2025 Update, December 2025).

For Work, we estimate \$25 billion to \$30 billion for metered observability and security event management, combining Gartner's observability-platform and security information and event management markets. We add \$1 billion to \$2 billion for agent runtime, our estimate informed by Menlo Ventures' December 2025 figures of \$750 million for agent platforms and \$1.5 billion for AI infrastructure in 2025.

Agents could expand these categories by making more work affordable and creating software that would never have been built by hand.

Knowledge work pays roughly \$25 trillion to \$35 trillion a year, our estimate from the 2.4 trillion knowledge-work hours in Appendix A5 at \$10 to \$15 an hour, against about \$58 trillion of all labor income (the ILO's labor income share of 52.3% applied to the IMF's 2024 world GDP). The sizing figure uses the upper end, which gives a sense of the economic activity involved. It is not a backend revenue forecast. Customers retain part of the benefit, useful work still requires review, and spending is divided among models, applications and infrastructure.

The financial model is narrower. It estimates the four case studies company by company. Their combined revenue reaches about \$71 billion in the Agentic Workforce and \$124 billion in the Machine Economy in 2032. Those totals include their existing businesses; modeled agent revenue contributes about \$15 billion and \$53 billion respectively.

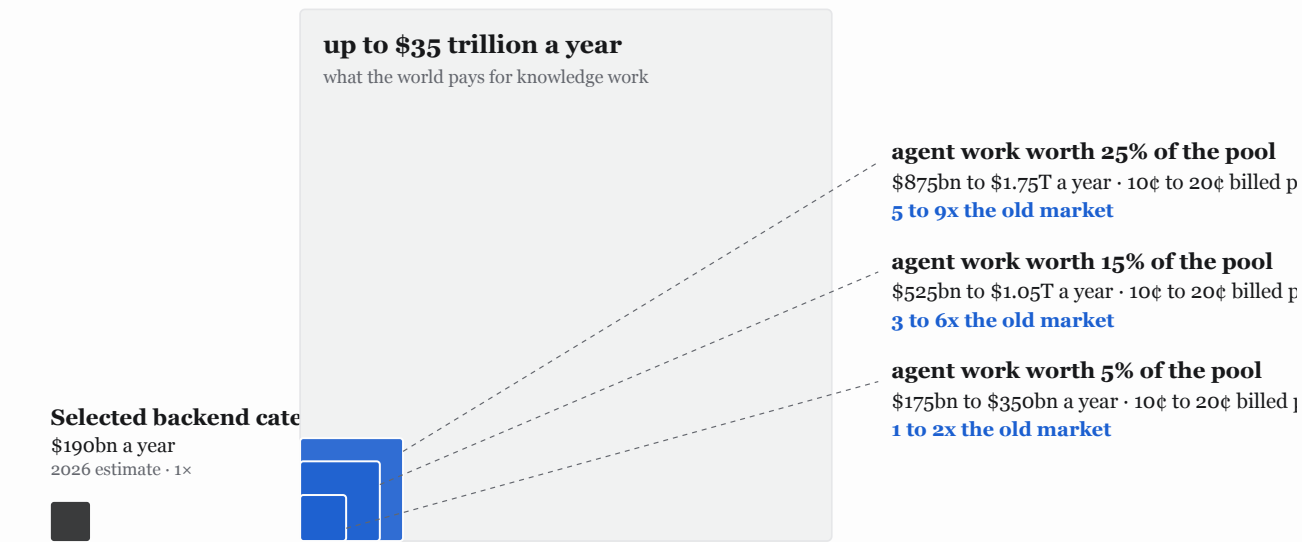
We believe the category can become much larger. The work ahead is to establish how much activity becomes durable paid use, and where that spending lands.

FIGURE 3.10

ILLUSTRATIVE

A large pool of work and a smaller infrastructure bill

The scale of knowledge-work wages beside current software spending and illustrative delegation scenarios



The illustration compares up to \$35 trillion of human labor value with about \$190 billion of backend spending, on the definitions in the text as of September 2026. Each tier assumes agents take on the stated share of that work and that providers bill 10 to 20 cents per dollar of human work delivered; both are illustrative assumptions, not estimates. These pools overlap with other economic activity and are not directly interchangeable. The comparison frames the scale of the question; it is not a forecast that labor spending transfers to infrastructure providers.

IV

How we will test the thesis

Financial scenarios, market expectations and the evidence that would change our view.

10 MIN READ · 9 FIGURES

How we will test the thesis What has the market rewarded so far? What do today's prices assume?

What must the businesses deliver? Can activity grow without much provider revenue?

What would change our view What evidence will settle the argument? How does the test differ by company?

Research commentary from [Skycatcher](#), LLC, an exempt reporting adviser; not an offer, solicitation, recommendation or investment advice. The index returns in this part are equal-weight baskets defined in 2026 and applied back to 2022: hypothetical, hindsight-selected and not investable. The revenue paths and valuation arithmetic for Cloudflare, MongoDB, Snowflake and Datadog are our estimates and illustrations of the scenarios, not company guidance and not price targets, and the companies have not reviewed them. [Skycatcher](#), its affiliates, principals and clients may hold long or short positions in any public company named here and may trade them without notice. [Appendix D3](#) carries the full disclosures.

How we will test the thesis

The machine-hours prediction and the financial scenarios answer different questions. The first asks how much agents work. The second asks what selected companies earn from that work. A rising share price does not verify either one.

We begin with what markets have rewarded and what current prices assume. We then examine what our revenue scenarios require, how they could fail and the evidence that would settle the argument. Finally, we apply those tests company by company. Share-price performance has its own grading rule; it cannot substitute for operating evidence.

What has the market rewarded so far?

The market has rewarded much of the hardware and data-center build ahead of the backend. Our research baskets show that sequence. They are hypothetical baskets assembled in 2026 and applied to earlier prices, so they contain hindsight and do not represent an investable track record.

The backend's average experience also hides large differences. Palantir materially changes the index result. Company size, revenue growth and valuation vary widely. The tracker shows those differences rather than asking one average to represent them all.

We compare price performance with reported revenue growth to see how market returns and business results differ.

FIGURE 4.1

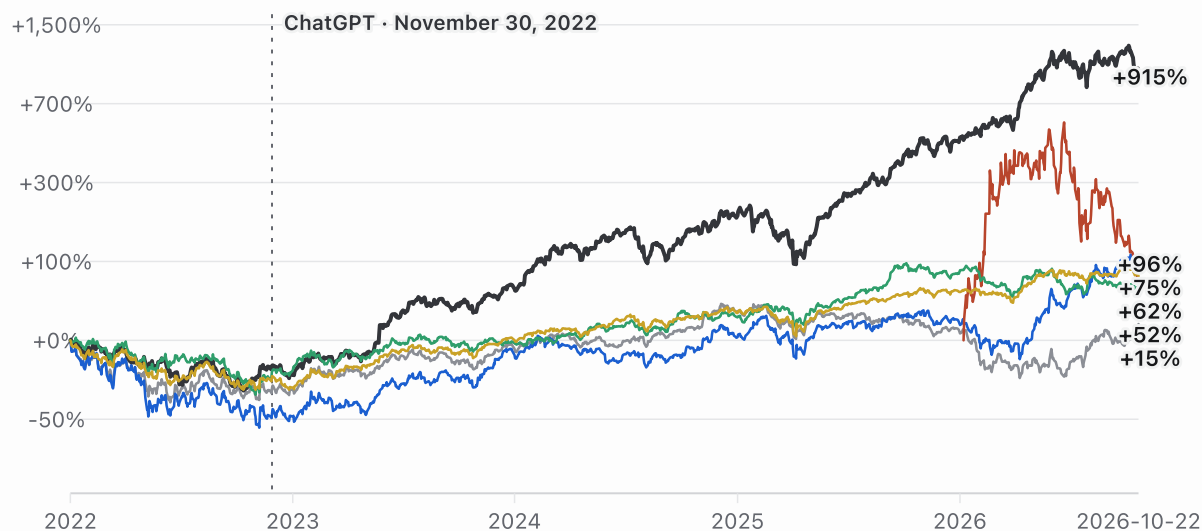
HYPOTHETICAL RESEARCH BASKETS · HISTORICAL PRICES, UPDATED DAILY

How the research baskets have traded

Pick the start date and the layers; every line is rebased to 0% at the start

EQUAL-WEIGHT INDICES · CHANGE SINCE JANUARY 3, 2022 · LOG SCALE · THROUGH OCTOBER 22, 2026

— Applications — Backend — Intelligence (from January 8, 2026) — Computation
— Hyperscalers — Nasdaq 100



[Follow the layers live on the tracker →](#)

Equal-weight, adjusted-price comparisons put the backend groups beside the Nasdaq 100 benchmark. The company roster was selected in 2026 and applied to history from January 2022, which introduces hindsight. The baskets are hypothetical and exclude fees, taxes and trading costs. Sources and method: [Appendix C1](#).

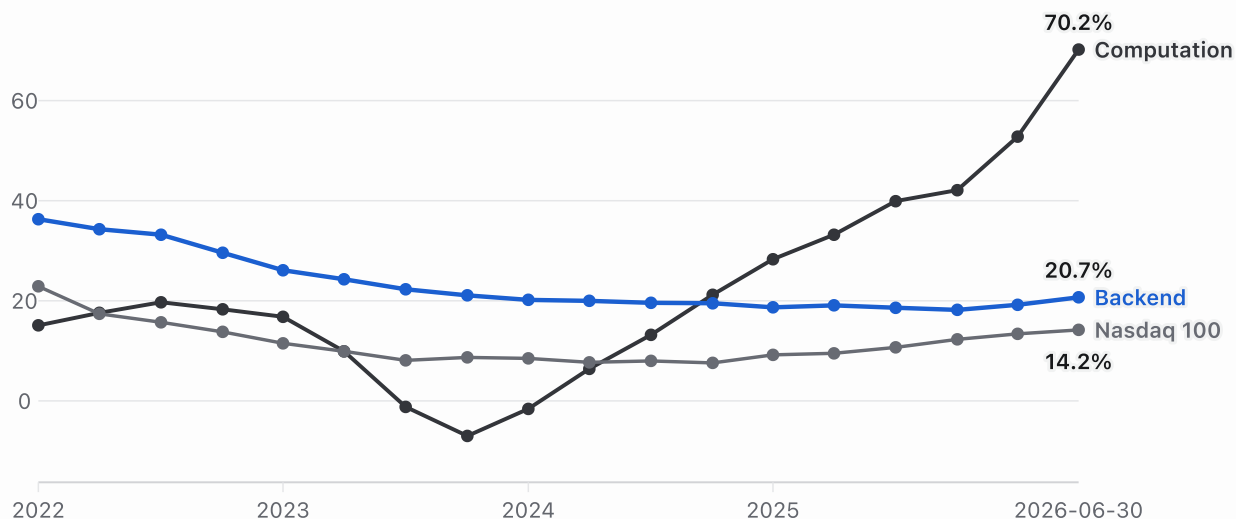
FIGURE 4.2

OBSERVED

Revenue growth across the backend

How fast revenue has grown: the backend against the Computation layer, quarter by quarter since 2022

MEDIAN REVENUE GROWTH, TRAILING TWELVE MONTHS, YEAR OVER YEAR, %



[Follow the two baskets' growth live on the tracker →](#)

The series compares trailing revenue growth for the covered companies. A quarter is shown only when at least 10 comparable companies report (50 for the Nasdaq 100), so coverage can differ from the full 32-company roster because not every company reports comparable quarterly figures. Company fiscal periods are not perfectly aligned. Sources and method: [Appendix C1](#).

What do today's prices assume?

The next question is what current prices require from the businesses. A revenue multiple compares a company's value with its revenue. Growth, profitability, dilution and risk all influence it.

Our valuation tool asks what revenue growth could support a given multiple under stated cash-flow assumptions. The standard case holds growth for five years, then fades it toward 3%, with a 25% free-cash-flow margin and a 10% discount rate.

On those assumptions, the backend's 6.9-times forward-revenue median at October 22, 2026, the latest close in the daily series, corresponds to roughly 16% annual growth for the first five years; the figure and this sentence update each trading day. Figure 4.3 shows the trailing-revenue multiple, a different basis. The four detailed case studies generally carried higher expectations. Each should be compared with its own scenario, on the same dates and revenue basis.

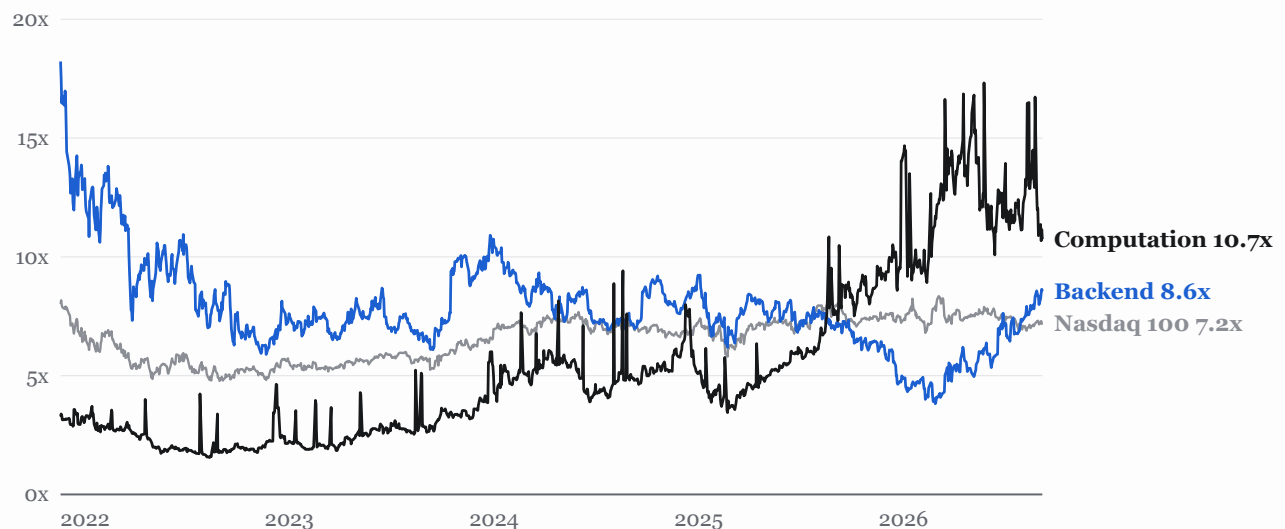
This is a sensitivity exercise. A multiple does not reveal a unique market forecast, and the result changes when margins or the discount rate change.

FIGURE 4.3

OBSERVED

What the market pays for current revenue

What the market pays for a dollar of revenue: the backend against the Computation layer since 2022



Backend, Computation and the Nasdaq 100: median enterprise value to trailing revenue across each basket, revalued every trading day. Computation is the 25-name layer of the stack roster, Appendix C1.

[Follow the multiples live on the tracker →](#)

Median enterprise value to trailing revenue provides a common valuation reference for each group. Enterprise values and revenue inputs retain their dates and coverage. The median describes the selected companies; it does not adjust for every difference in margins, growth, dilution or business quality.

Growth consistent with the latest forward multiple

Choose a revenue growth rate and compare the resulting illustrative multiple with the market's multiple

Revenue growth a year, for five years  **28%**

Set it to

Latest close

Agentic Workforce 28%

Machine Economy 38%

Free-cash-flow margin

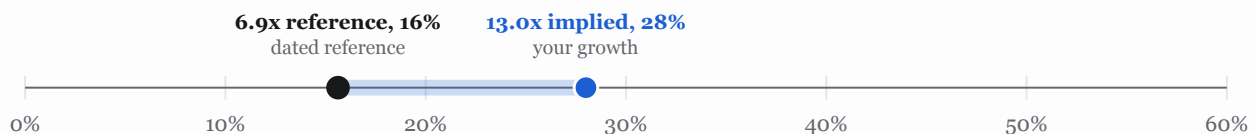
Margin held at 25%

Margin rising to 35%

The observed multiple is the backend basket's forward-revenue median at the latest close in the daily series, October 22, 2026; the tracker carries the same reading each trading day. The scenario buttons borrow annual growth rates from four companies over 2025–2032 and apply them for five years here. These are different groups and periods: use them to explore the assumptions, not to infer a valuation gap.

13.0x forward revenue under 28% growth and the selected cash-flow assumptions

The backend's median was **6.9x** on October 22, 2026. Under these assumptions, that corresponds to **16%** annual growth for five years. The calculated multiple is **1.9 times** the dated reference.



Growth implied by the valuation assumptions: revenue grows at the chosen rate for five years, then fades to 3% by year 11; the free-cash-flow margin starts at 25% and follows the path chosen; cash flows discounted at 10%. The multiple is enterprise value over next-twelve-month revenue. On these terms 27.7% annual growth corresponds to 12.8x and 38.3% about 22x at a 25% margin. The case you build with the slider is your own: not a [Skycatcher](#) estimate, a forecast or a recommendation.

The reverse valuation calculation asks what revenue path supports the observed market value under the chosen margin, discount-rate and terminal-value assumptions. A different assumption changes the required growth. The result is a conditional calculation, not a market forecast or price target.

What must the businesses deliver?

Our four company cases make the revenue assumptions explicit. Together, the companies reported about \$12.7 billion of combined revenue in the model's 2025 base. Our scenarios reach approximately \$40 billion to \$49 billion in 2029, against the analyst-consensus snapshot of \$32.7 billion registered on September 8, 2026; the October 7, 2026 consensus is \$33.1 billion.

Most of that difference comes from the growth we assume for the existing businesses. Modeled agent revenue is only about 2% to 3% of the 2029 totals. The near-term forecast therefore depends on reported growth lasting longer or accelerating more than the consensus snapshot assumed.

Commitments and customer expansion offer evidence, but they do not secure several years of growth. We watch consumption, renewals and reported revenue each quarter. A renewed period of customer cost-cutting would weaken this part of the forecast long before the longer-term hours question is settled.

After 2029, the modeled agent contribution becomes much larger. By 2032 it is about a fifth of combined revenue in the Agentic Workforce and more than two fifths in the Machine Economy. That is when the financial outcome depends more heavily on agent hours arriving and reaching these providers.

FIGURE 4.5

OBSERVED · REPORTED, AND SCENARIO

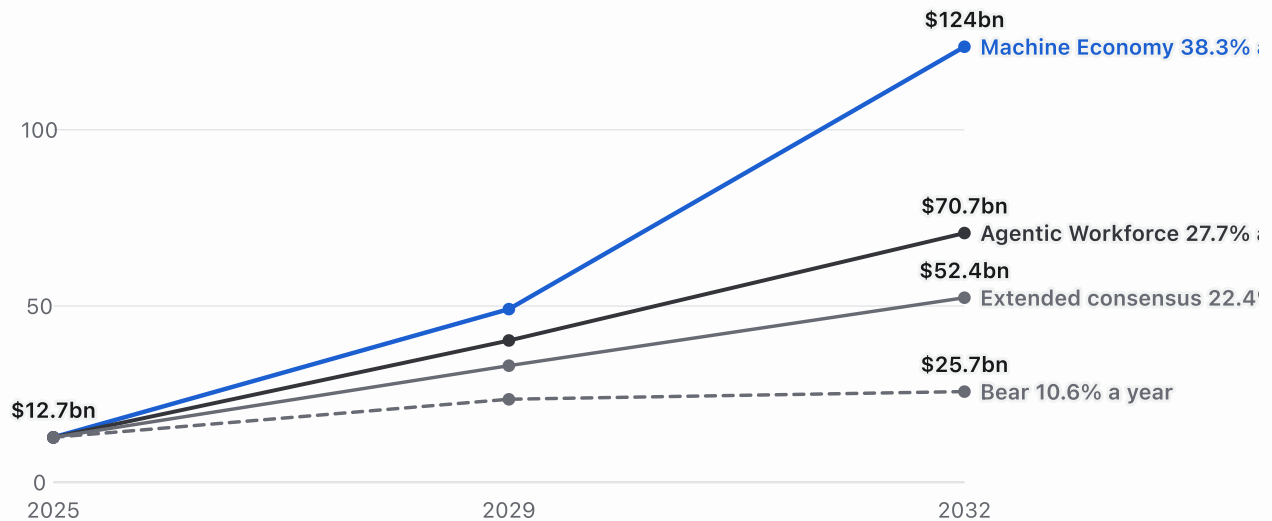
Four revenue paths with different foundations

Combined revenue of the backend bellwethers under four cases, 2025 to 2032

Both scenarios run ahead of consensus from 2027 on the reported rates. The Machine Economy goes further.

From 2022 to 2025 the four compounded revenue at 28.5% a year. The four paths from here:

COMBINED REVENUE OF THE FOUR COMPANY CASES, \$BN



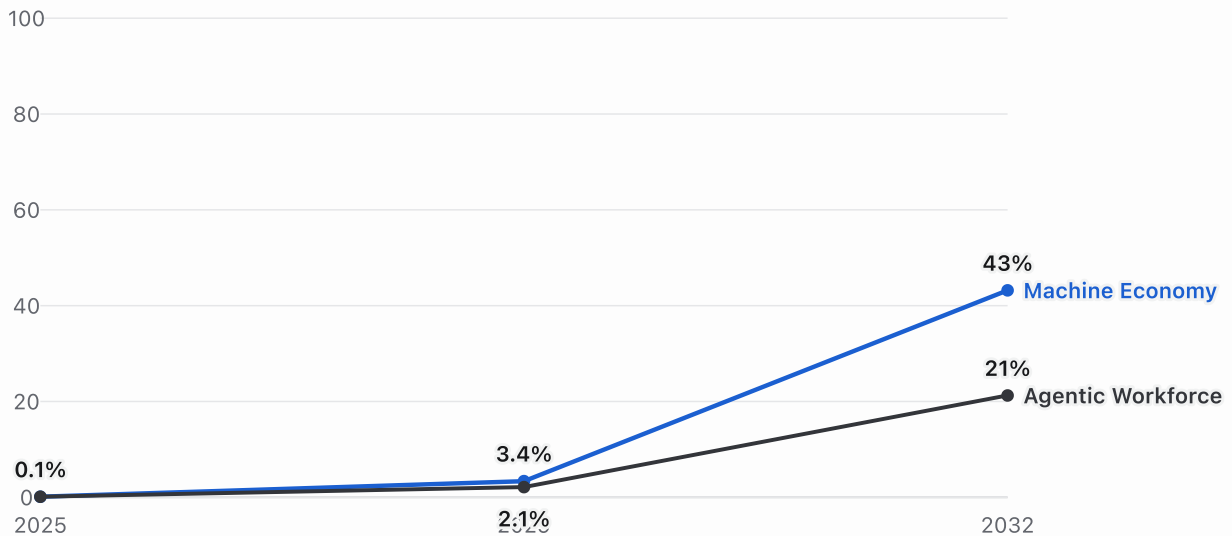
[Open the revenue model behind this figure →](#)

Reported revenue, dated analyst consensus and the two [Skycatcher](#) scenarios are shown separately. The consensus extension after 2029 is our calculation, not a published analyst estimate. Each [Skycatcher](#) path combines existing-business growth with separately modeled agent revenue. The four companies do not represent the whole backend opportunity.

When agent revenue becomes material in the model

How much of the four companies' revenue comes from agents: under 1% in 2025, about 2.1% to 3.4% by 2029 and 21% to 43% by 2032

AGENT-ATTRIBUTABLE SHARE OF COMBINED REVENUE, %



[Open the revenue model behind this figure →](#)

The separately modeled agent line contributes about 2.1% and 3.4% of combined revenue in 2029, rising to about 21.3% and 43.2% in 2032. Those shares belong to the Agentic Workforce and Machine Economy respectively. They are model outputs, not company-reported agent revenue. Existing-business revenue already includes some AI use. Sources and method: [revenue model](#).

Can activity grow without much provider revenue?

Those revenue paths require more than agent adoption. The strongest commercial countercase accepts the machine-hours forecast. Agents work much more, but open-source systems, cloud bundles and internal software serve most of the growth. Prices also fall faster than our central assumptions.

We model that combination explicitly. With Machine Economy hours, flat provider share, steeper price declines and slower growth in the existing businesses, combined revenue is about \$59 billion in 2032. Agent activity is enormous, while the modeled agent revenue at the four companies is only about \$3 billion.

This is why the category view cannot be reduced to a claim about four stocks. New companies can succeed while established providers lose share. Customers may capture most of the economic benefit.

What would change our view

The risks span the whole chain, from repeat use to the cash providers keep.

- 1. **Repeat use remains low.** Agents can finish impressive demonstrations but are rarely trusted with recurring work.
- 2. **Review remains too expensive.** A low inference bill does not produce a low cost per accepted result.
- 3. **Capacity falls short.** Power, hardware or efficiency improvements cannot support the hours demanded.
- 4. **New workloads choose other systems.** Global activity rises while the providers we follow serve a smaller share.
- 5. **Prices fall faster than paid use grows.** Customers do more work while spending less.
- 6. **Growth fails to become cash.** Model costs, operating expenses or dilution absorb the revenue increase.
- 7. **The economic setting changes fundamentally.** Our financial scenarios assume a functioning market economy and useful demand for the work performed.

FIGURE 4.7

STRUCTURE

What could weaken the thesis

Seven ways the thesis fails, what we say to each, and what we lean on

RISK	OUR COUNTERPOINT, AND THE EVIDENCE
1 Repeat use remains low.	<p>Agents can finish impressive demonstrations but are rarely trusted with recurring work. Most near-term revenue upside comes from assumptions about the existing businesses, and aggregate growth alone cannot prove agents caused it. We need evidence that additional agent activity becomes paid use.</p> <p>Evidence to watch: both current hours estimates flat for two quarters against paths that need about 2.9x a year; paid conversion falling on the platforms that disclose it; revenue per completed task shrinking. An illustrative slower-growth path reaches \$25.7 billion of 2032 revenue; the model page’s constrained-adoption case is \$47 billion.</p>
2 Review remains too expensive.	<p>A low inference bill does not produce a low cost per accepted result. Reliable execution, permission checks and audit records can limit the harm from an agent’s mistakes, but they cannot make every task useful; the thesis needs acceptable results at a cost customers will keep paying.</p> <p>Our coding logs show about 52 minutes between turns per active hour (Appendix B1). We have not measured how much of that time is human review. Capability is moving: METR’s 50%-success task horizon reached about 12 hours on February 5, 2026, before METR stopped updating this series in September. The tracker preserves the historical record; we assess newer capability evidence separately.</p>
3 Capacity falls short.	<p>Power, hardware or efficiency improvements may not support the hours demanded. The capacity path in Appendix A3 allows roughly four trillion hours in 2029 and, with eightfold compute and tenfold efficiency, about 357 trillion by 2032; the Machine Economy needs 340. The model shows demanded and delivered hours separately.</p> <p>Evidence to watch: announced campuses not energized on schedule; disclosed inference capacity below the path; delivered hours falling below demand on the model page.</p>

4 New workloads choose other systems.

They can. Postgres, Databricks, open-source and in-house stacks already take new workloads, a model provider can run its own databases, queues and monitoring, and bundled cloud services remain a direct competitor. Neon and Supabase report agents creating most new databases; those disclosures measure creation, not paid retention. We track which services remain useful and paid for as the application grows.

AWS, Google, Microsoft and Databricks all shipped agent memory in 2026, one priced at 25¢ per thousand memories. Evidence to watch: a frontier lab offering state, execution or observability as a paid product to third parties; the four's share of hours falling while machine hours rise (the tracker's provider-use series); AI-native customers reducing use at renewal, as Datadog's largest did in 2026.

5 Prices fall faster than paid use grows.

The model assumes lower unit prices as volume rises: 50% to 65% of the previous price after each tenfold increase, depending on the company and scenario. Customers could do more work while spending less.

Token use rose about 2,000x from 2022 to 2026 while the price of a token fell about 200x, and the labs spent the gains on reasoning models that think for minutes. The thesis holds while tasks multiply faster than the cost of each falls.

6 Growth fails to become cash.

Model costs, operating expenses or dilution can absorb the revenue increase. The revenue model prices hours, not profit; the valuation sensitivity in Figure 4.4 assumes a 25% free-cash-flow margin, and the downside record in Appendix D1 reads cash conversion over two trailing four-quarter periods.

Evidence to watch: free-cash-flow margins falling while revenue grows; serving costs or share dilution absorbing the increase; the four's cash conversion deteriorating (the tracker's valuation section).

7 The economic setting changes fundamentally.

Our financial scenarios assume a functioning market economy in which customers pay for useful services. A more radical change would require a different framework.

We follow measured capability and deployment. We do not assign revenue to an unmeasured leap in intelligence.

The seven risks above, each paired with our counterpoint and the observation that would test it. A larger activity total is insufficient if work stays unreliable or providers fail to earn the expected revenue.

What evidence will settle the argument?

We turn these questions into dated predictions. The headline hours prediction sits alongside specific tests of capability, adoption, company revenue and market performance. Each has a baseline, a measure and a deadline.

The next readings concern adoption and company growth. We will publish outcomes against the original launch terms, whether predictions succeed or fail. A later forecast revision does not erase the original call.

We believe machine work can become a major source of infrastructure demand. The tracker is there to establish how quickly that is happening and who, if anyone, is earning durable revenue from it.

Predictions we will grade

The headline prediction · registered October 8, 2026 · graded at the 2032 reading, by January 31, 2033

PREDICTION	DEADLINE	STATUS
H1 · Machine hours exceed all paid human work, 7.0 trillion hours a year, by the 2032 reading.	January 31, 2033	Open
▼ HOW IT IS MEASURED AND ADJUDICATED		
<p>Observe the last complete week of December 2032; multiply daily active agent hours by 365. Compare with the fixed 2026 paid-human-work benchmark: central 7.0 trillion annual hours, range 6.5–7.5 trillion. A robust hit requires both the platform level and OpenRouter pulse lower sensitivity bounds above 7.5 trillion. A miss requires both upper bounds below 6.5 trillion. All overlap, disagreement, missing evidence or failed method review is unresolved. Publish the two central estimates as well as ranges; these are assumption ranges, not statistical confidence intervals. Later evidence can restate that week, never count activity after it. Initial grade January 31, 2033; one final restatement by June 30, 2033 may remain unresolved. An independent methodological review must be published by September 30, 2032. A replacement measured panel and its uncertainty rule may govern only if registered with that review at least a quarter in advance. Method changes are recorded under Appendix A4 and do not count as activity growth. The claim concerns active duration, not equivalent output or displaced jobs.</p>		

Next 12 months · by October 1, 2027

BET	DEADLINE	STATUS
B1 · The backend index beats the Computation layer over the 12 months from October 1, 2026.	October 1, 2027	Open
▼ HOW IT IS MEASURED		
<p>The equal-weight 32-name backend index against the 25-name Computation layer basket, both rebased to 0% at the September 30, 2026 close, the last close before October 1, 2026, and read at the October 1, 2027 close on the tracker’s index chart, start set to Since October 1, 2026. Both baskets are frozen at the launch roster; the roster rule below governs every later change. Launch roster rule: membership is frozen in launch-rosters.json. No new IPO or discretionary addition enters. Equal-weight daily returns use the same adjusted-close method as the published index. A delisting uses the last available adjusted close, then that position stays in cash at a zero return until the next trading day, when surviving members are equally reweighted; no replacement enters. If reliable delisting proceeds are unavailable, publish the missing treatment and leave the grade unresolved. Splits and dividends follow the adjusted-close series; acquisitions never add a new name. The launch-cohort grade must be published separately if the live tracker roster later changes.</p>		
B2 · Median backend revenue growth improves at least three points from the quarter ended June 30, 2026.	September 30, 2027	Open
▼ HOW IT IS MEASURED		
<p>Median year-over-year revenue growth of the matched panel: the backend companies that report a comparable quarterly figure for both the baseline quarter and the test quarter (at the baseline, 30 of 32; Megaport reports half-yearly and is excluded). Baseline 22.65% for the quarter ended June 30, 2026, so the threshold is 25.65%. The observation period is the quarter ended June 30, 2027 (or each company’s last quarter ending on or before that date); the grade is read by September 30, 2027, allowing time for filings. Missing or incomparable test-period filings leave the bet unresolved; companies cannot be dropped after the fact to improve the median.</p>		
B3 · At least 17 of the 32 backend companies disclose an aggregate AI-attributable revenue, ARR, run-rate or bookings amount, or a share of revenue.	September 30, 2027	Open
▼ HOW IT IS MEASURED		
<p>Count each company in the frozen launch roster once if its last four reported quarters as of September 30, 2027 contain a company release, filing or company-hosted transcript with an aggregate AI-attributable revenue, ARR, run-rate or bookings amount, or a share of revenue. AI-native customer cohorts count when the company quantifies their revenue or ARR, but are separately identified: their whole bill is not necessarily AI workload. Growth rates, shares of growth, product adoption, customer counts, individual contract values and undifferentiated platform-wide consumption do not qualify. All 32 use this same rule. The target is 17; the source register distinguishes qualifying, excluded and unverified records. Unknown records remain unverified, not zero. Publish the underlying source and metric for every company counted at grading.</p>		

BET	DEADLINE	STATUS
B4 · A publicly evaluated model reaches an 80% task horizon of at least eight human-expert hours. ▼ HOW IT IS MEASURED METR's 80%-success horizon, central estimate, threshold 480 minutes. The horizon is the task duration, in human-expert time, at which METR's fitted curve gives the model an 80% chance of success; it is neither the longest task the model completed nor elapsed agent runtime. The grade reads the central estimate as METR publishes it at the deadline; a later change in METR's method does not reopen it. METR said on September 8, 2026 that its Time Horizon page is no longer actively updated; if no newer central estimate exists at the deadline, the grade reads the last published one and says so.	June 30, 2027	Open

B5 · OpenRouter's daily tokens triple. ▼ HOW IT IS MEASURED Complete-week average daily tokens against the registered baseline of 16.495 trillion a day (week marked September 6, 2026); threshold 49.49 trillion. Token volume is a usage indicator; it is not agent hours or spending.	September 8, 2027	Open
---	-------------------	------

B11 · The four bellwethers' combined revenue growth stays at or above 30% year over year in each of the next four reported quarters. ▼ HOW IT IS MEASURED Summed latest-quarter revenue of Cloudflare, MongoDB, Snowflake and Datadog against the same four fiscal quarters a year earlier, from SEC filings, for the four quarters reported after October 1, 2026: for Cloudflare and Datadog the quarters ending September 30 and December 31, 2026 and March 31 and June 30, 2027; for Snowflake and MongoDB the quarters ending October 31, 2026 and January 31, April 30 and July 31, 2027. Baseline 34.5% on the quarters ended June 30, 2026 (Cloudflare, Datadog) and July 31, 2026 (Snowflake, MongoDB). One quarter below 30% fails the bet.	September 30, 2027	Open
---	--------------------	------

Beyond 12 months · grading dates from January 2028 to March 2032

BET	DEADLINE	STATUS
B7 · AI agents performing user-directed work exceed human-originated requests in the disclosed traffic of at least two major platforms. ▼ HOW IT IS MEASURED A platform's own disclosure of its traffic by origin, with the same denominator in both readings (all requests the platform serves): requests from AI agents acting on a person's instruction, as the platform classifies them, against requests from people; ordinary bots, crawlers and scrapers are excluded from the agent share. A major platform is one of Cloudflare, GitHub, OpenRouter, Vercel and Stripe, or any platform disclosing more than one trillion requests a month; two of them must disclose the agent share above the human share. A research classification of a platform's volume, including ours, does not qualify.	December 31, 2031	Open

B8 · The 32-name backend reports more than \$250 billion of annual recognized revenue. ▼ HOW IT IS MEASURED Sum recognized revenue for fiscal years ending in calendar 2031 for the frozen 32-name launch roster, from company filings; this is a fiscal-year sum, not calendarized revenue. Translate Megaport at the average exchange rate for its fiscal year and publish the rate. Count a business only once: if acquired by another constituent, include it through the consolidated acquirer and add only the target's pre-consolidation revenue falling in the acquirer's reporting period, where separately disclosed. A target acquired outside the roster contributes only separately disclosed revenue for the relevant fiscal period; never carry its last four quarters forward as if they were 2031 revenue. Undisclosed, missing or incomparable amounts leave the grade unresolved. Baseline \$70.92 billion is the published 2025 fiscal-aligned sum. Threshold and March 31, 2032 grading deadline unchanged.	March 31, 2032	Open
---	----------------	------

B12 · Cloudflare, MongoDB, Snowflake and Datadog report more than \$40 billion of combined revenue for the 2029-aligned reporting periods, against the registered \$32.7 billion consensus snapshot. ▼ HOW IT IS MEASURED Reported revenue for calendar 2029 at Cloudflare and Datadog and fiscal years ending January 31, 2030 at Snowflake and MongoDB. Threshold: \$40 billion combined. The registered consensus comparator is \$32.7 billion from the September 8, 2026 snapshot; current consensus is displayed separately and does not reset it. Initial and final grade March 31, 2030. Missing filings leave the grade unresolved.	March 31, 2030	Open
--	----------------	------

B14 · The backend is the best-performing layer of the stack over the three years from October 1, 2026.

October 1, 2029 Open

▼ HOW IT IS MEASURED

The same lines read at the October 1, 2029 close, from the September 30, 2026 close in the published index file, the start labelled Since October 1, 2026; the backend must show the largest gain. Launch roster rule: membership is frozen in launch-rosters.json. No new IPO or discretionary addition enters. Equal-weight daily returns use the same adjusted-close method as the published index. A delisting uses the last available adjusted close, then that position stays in cash at a zero return until the next trading day, when surviving members are equally reweighted; no replacement enters. If reliable delisting proceeds are unavailable, publish the missing treatment and leave the grade unresolved. Splits and dividends follow the adjusted-close series; acquisitions never add a new name. The launch-cohort grade must be published separately if the live tracker roster later changes. Graded separately as a market call; it overlaps with B1 and does not count as independent confirmation of the operating thesis.

B15 · By the end of 2027, AI agents perform more than 100 million active hours a day worldwide.

January 31, 2028 Open

▼ HOW IT IS MEASURED

Observe the last complete week of December 2027, fixed even if evidence arrives later. Threshold: 100 million active agent hours a day worldwide, equivalent to 36.5 billion annualized hours. Read both the platform level and OpenRouter pulse under the dated Appendix A4 method. Publish both central estimates and assumption ranges. A robust hit requires both lower bounds above the threshold; a miss requires both upper bounds below it. Overlap, disagreement or missing evidence is unresolved. These are sensitivity ranges, not statistical confidence intervals. No single leg overrules the other at final grading. Initial grade January 31, 2028; one restatement of that same week by June 30, 2028 is final, including an unresolved grade. Independently review and register the method and range construction by September 30, 2027. This is active duration, not equivalent human output or jobs displaced.

Dated October 8, 2026. The launch slate, October 8, 2026: eleven bets with stated measures and grading dates. H1 is the headline prediction. A test of paid retention of agent-built software waits on a matched 90-day baseline; no numerical target is set until that evidence exists. A bet is graded once at its deadline and the grade is published beside it; the registered predictions are recorded in Appendix D2. Beyond 2031, three calls stay on the record: A usage-oriented infrastructure business exceeds \$100 billion of annual recognized revenue by December 31, 2036; The public backend universe reaches \$1 trillion of annual recognized revenue by December 31, 2038; An external index provider launches an agent-infrastructure index by December 31, 2040.

The register contains eleven supporting predictions with stated measures and deadlines, including a machine-hour milestone for the end of 2027. H1 separately records the main prediction that machine hours pass the human benchmark by the end of 2032. Original terms remain fixed. Outcomes are published whether the prediction succeeds or fails. Sources and method: [Appendix D2](#).

How does the test differ by company?

The same tests can produce different answers for each provider. The company cards bring product exposure, reported evidence, scenario outcomes and the strongest countercase together. We judge each business on the paid work it serves and the evidence it reports.

Company cases as tests of the framework

Four company cases: the workload, reported adoption and commercial evidence, reasons customers might stay, market expectations and evidence that would weaken the case

Cloudflare

THE WORKLOAD IT COULD WIN

Running agent software, handling its requests and keeping track of its work. Multistep tasks can generate repeated requests; the volume depends on the task.

REPORTED ADOPTION AND COMMERCIAL EVIDENCE

Activity: Workers developers rose from 3 million in December 2024 to 7.4 million in June 2026; agent requests rose more than 1,700% in the year to May 2026. Contract value: an \$8.7 million, two-year agreement with a coding-app company in Q1 2026. No separately reported AI-attributable revenue.

WHY THE WORKLOAD WOULD STAY

Moving an application can require substantial changes when it depends on platform-specific services. Large cloud providers could win more enterprise workloads, leaving Cloudflare with smaller customers.

VALUATION UNDER STATED ASSUMPTIONS

45.8x trailing revenue corresponds to about **53%** growth a year for five years on Figure 4.4's terms; our scenarios carry **30%** and **42%** for this company over 2026 to 2031 (enterprise value as of 2026-10-23).

THE NEXT OBSERVATION THAT WOULD WEAKEN THE CASE

A quarter in which developer growth continues while revenue growth decelerates would weaken the case; it is consistent with the free tier absorbing the agents, but also with pricing, mix and timing. The next check: paid Workers accounts and revenue per developer customer group, which Cloudflare has not disclosed.

Sources: Investor Day · Jun 2026 · Q1 2026 call · Q2 2026 transcript (secondary)

MongoDB

THE WORKLOAD IT COULD WIN

Storing application data and agent memory, and helping agents retrieve relevant information. The opportunity depends on how many new applications reach sustained use.

REPORTED ADOPTION AND COMMERCIAL EVIDENCE

Customers: AI-native Atlas customers rose from 1,200 to 2,200 from Q2 FY25 to Q2 FY27. Customer revenue: up 165% year over year; this group remains a low-single-digit share of revenue. Customers with any AI use case account for 40% of Atlas annual recurring revenue in Q2 FY27. Activity: Managed MCP Server use grew 8x from its first week.

WHY THE WORKLOAD WOULD STAY

Stored data and existing connections can make a database costly to replace. The test is whether new applications keep running and become paying customers.

VALUATION UNDER STATED ASSUMPTIONS

9.9x trailing revenue corresponds to about **21%** growth a year for five years on Figure 4.4's terms; our scenarios carry **24%** and **33%** for this company over 2026 to 2031 (enterprise value as of 2026-10-23).

THE NEXT OBSERVATION THAT WOULD WEAKEN THE CASE

Atlas consumption growth slowing while AI-native customer counts rise would weaken the case; it is consistent with applications created and abandoned, but also with optimization and customer mix. The next check: paid conversion of agent-created databases and consumption per active AI-native customer.

Sources: [Investor Day](#) · [Sep 2026](#) · [Q2 FY27 call](#)

Snowflake

THE WORKLOAD IT COULD WIN

Helping agents query company data under existing access controls. Our model assumes the highest bill per hour among these providers, but the smallest share of machine hours.

REPORTED ADOPTION AND COMMERCIAL EVIDENCE

Activity: 9,100 accounts used Cortex Code in the quarter to July 31, 2026, with 2,000 added; 5,800 used CoWork. Management says Cortex Code accounts consume more existing services. Annualized revenue: a \$100 million AI run rate disclosed in December 2025, not \$100 million recognized in that period. A consistent series is not available.

WHY THE WORKLOAD WOULD STAY

Keeping data in place can reduce transfer and integration costs. The risk is that customers move more work to other databases or data platforms.

VALUATION UNDER STATED ASSUMPTIONS

21.1x trailing revenue corresponds to about **36%** growth a year for five years on Figure 4.4's terms; our scenarios carry **29%** and **40%** for this company over 2026 to 2031 (enterprise value as of 2026-10-23).

THE NEXT OBSERVATION THAT WOULD WEAKEN THE CASE

Cortex Code accounts growing while product revenue growth does not re-accelerate would weaken the case; it is consistent with agents reading the data without paying for compute, but also with contract timing and credit optimization. The next check: AI Credits consumed and the consumption of accounts that adopted Cortex Code, by customer group.

Sources: Q2 FY27 call · Q3 FY26 call

Datadog

THE WORKLOAD IT COULD WIN

Monitoring the software agents use, recording their actions and investigating failures. More activity can increase demand; the bill depends on how much data customers collect and retain.

REPORTED ADOPTION AND COMMERCIAL EVIDENCE

Customer revenue: AI-native customers supplied 12% in Q3 2025, from 6% a year earlier, the latest disclosed share. Activity: in Q1 2026, 6,500 customers sent AI data; records of model activity nearly tripled quarter over quarter. On August 6, 2026, Datadog reported lower usage from its largest customer and included the reduction in its guidance.

WHY THE WORKLOAD WOULD STAY

Monitoring can be hard to replace once teams rely on its alerts and response procedures. Customers can still cut spending by collecting less data or using monitoring built into other platforms. Dependence on a few large AI customers adds risk.

VALUATION UNDER STATED ASSUMPTIONS

23.4x trailing revenue corresponds to about **38%** growth a year for five years on Figure 4.4's terms; our scenarios carry **28%** and **39%** for this company over 2026 to 2031 (enterprise value as of 2026-10-23).

THE NEXT OBSERVATION THAT WOULD WEAKEN THE CASE

More monitoring activity without faster revenue growth from AI-native customers would weaken the case. Lower data retention and large-customer renewals could also explain the gap. Watch revenue from this customer group and billed activity per customer; these are not consistently disclosed.

Sources: [Q3 2025 call](#) · [Q1 2026 call](#) · [Q2 2026 call](#)

A growing backend market does not guarantee success for each company or a return at any price. Adoption entries use dated company disclosures. The valuation comparison uses the published enterprise value and the stated model assumptions.

The cases connect product exposure, reported financial evidence and conditional model outcomes. They illustrate how to investigate the revenue providers earn from the work. They do not establish a recommendation, a target price or that rising agent activity will favor every company equally.

The four cases illustrate the method. The wider opportunity includes new providers, open-source systems and cloud platforms. Part Five asks what they are building for agents and what a shared operating layer could make possible.

V

Building for agents

New companies, simpler backends and the longer-range vision of an operating layer for the machine economy.

13 MIN READ · 13 FIGURES

[Building for agents](#)

[Where new companies are forming](#)

[Funding the backend build-out](#)

[Fewer systems to connect](#)

[What adoption can establish](#)

[From assembling services to running software](#)

[What we have learned by building](#)

[An operating layer for the machine economy](#)

[What would make that vision real](#)

Research commentary from [Skycatcher](#), LLC, an exempt reporting adviser; not an offer, solicitation, recommendation or investment advice. The private-company valuations in this part are the companies' own announced round prices, not our estimates of value. [Skycatcher](#) is an investor in Clockwork Labs, the company behind Spacetime (SpacetimeDB), discussed below, and this site runs on its Maincloud service. [Skycatcher](#), its affiliates, principals and clients may hold positions in any company named here and may trade them without notice. [Appendix D3](#) carries the full disclosures.

Building for agents

Agents are becoming builders and users of software. They can choose a database, deploy an application and keep it running. The infrastructure beneath them has to become easier to use and dependable enough to carry the work.

We see a broad opportunity here. Established clouds, open-source projects and new companies are approaching the same problem from different directions. Some will sell individual services. Others will bring more of the system together. The architecture that serves machine work may look different from the one built around human developers.

Where new companies are forming

Our private-company map follows twenty-eight businesses across four groups: services that run agents, clouds that serve models, tools that control and monitor activity, and systems that store data and memory.

Temporal, Modal and Browserbase help run work. Fireworks, Baseten and Together serve models. Grafana, Braintrust and WorkOS help monitor or control activity. Databricks, Supabase, Convex and Spacetime supply data or application infrastructure. Their products overlap, but each gives us a different view of what agents need.

The first signs appear in creation. Vercel reported on June 17, 2026 that coding agents trigger more than half of its deployments. Databricks, which owns Neon, reported in January 2026 that agents create 80% of the databases on Neon. Supabase reported in June 2026 that AI tools launched more than 60% of its new databases, and in October 2026 that the share had reached 70%.

Those disclosures count different things over different periods. A database may last seconds. A deployment may never attract a user. We want to know how many of these applications remain useful, active and paid for 90 days after creation. Creation tells us where agents are arriving. Retention tells us whether the work lasts.

The private backend landscape

The private backend in four layers, with the latest priced valuation

Runtime	Inference cloud	Control plane	State
Execution	Model time	Watch and authorize	Read and write
Cognition \$48bn valuation · 2026	Fireworks AI \$18bn valuation · 2026	Grafana Labs \$9.0bn valuation · 2026	Databricks \$190bn valuation · 2026
Lovable \$13bn valuation · 2026	Baseten \$13bn valuation · 2026	Vanta \$4.2bn valuation · 2025	ClickHouse \$15bn valuation · 2026
Temporal \$13bn valuation · 2026	Together AI \$8.3bn valuation · 2026	WorkOS \$2.0bn valuation · 2026	Supabase \$10bn valuation · 2026
Vercel \$9.3bn valuation · 2025	fal \$4.5bn valuation · 2025	PostHog \$1.4bn valuation · 2025	Redis \$2.0bn valuation · 2021
Replit \$9.0bn valuation · 2026		Braintrust \$800m valuation · 2026	Neo4j \$2.0bn valuation · 2024
Modal \$4.7bn valuation · 2026		Clerk Series C \$50m raised · 2025	Pinecone \$750m valuation · 2023
n8n \$2.5bn valuation · 2025			Convex Series B \$57m raised · 2026
LangChain \$1.3bn valuation · 2025			Spacetime Series A \$22m raised · 2022
Browserbase \$300m valuation · 2025			
E2B Series A \$21m raised · 2025			

Twenty-eight private companies show the range of products and architectures being built. Funding and valuation entries use the latest cited transactions or disclosures, with dates. A last-round valuation is a historical transaction reference, not a live market value or proof of operating performance. Cursor left the map when SpaceX’s acquisition closed (8-K filed August 14, 2026). Sources and method: Appendix C3, which carries each company’s dated source.

What adoption disclosures actually count

The share of usage that agents now account for, as the companies disclosed it

Vercel · deployments triggered by coding agents

more than 50%



under 3% on December 17, 2025 · June 17, 2026 · [source](#)

Neon (Databricks) · databases created by agents

80%



January 27, 2026 · [source](#)

Supabase · new databases launched by an AI tool

70%



more than 60% on June 4, 2026 · October 2, 2026 · [source](#)

Vercel · applications on the platform that originated from agents

30%



April 13, 2026 · [source](#)

Developer registrations, database creation, projects, deployments and revenue measure different things. Each disclosure carries the company's original definition, date and population. Product creation can be an early signal; paid retention and repeat production use provide stronger evidence of a lasting business.

Funding the backend build-out

The companies in our map raised much more capital from 2024 to 2026 than in the two years before. We record primary equity separately from debt and secondary share sales. Announced round prices show what investors paid at a point in time.

Funding can pay for capacity and distribution. The business still has to earn repeat use. We expect demand to grow for agent memory, identity, permissions, monitoring and application hosting. Some of that spending will reach new providers; some will be included in products customers already buy.

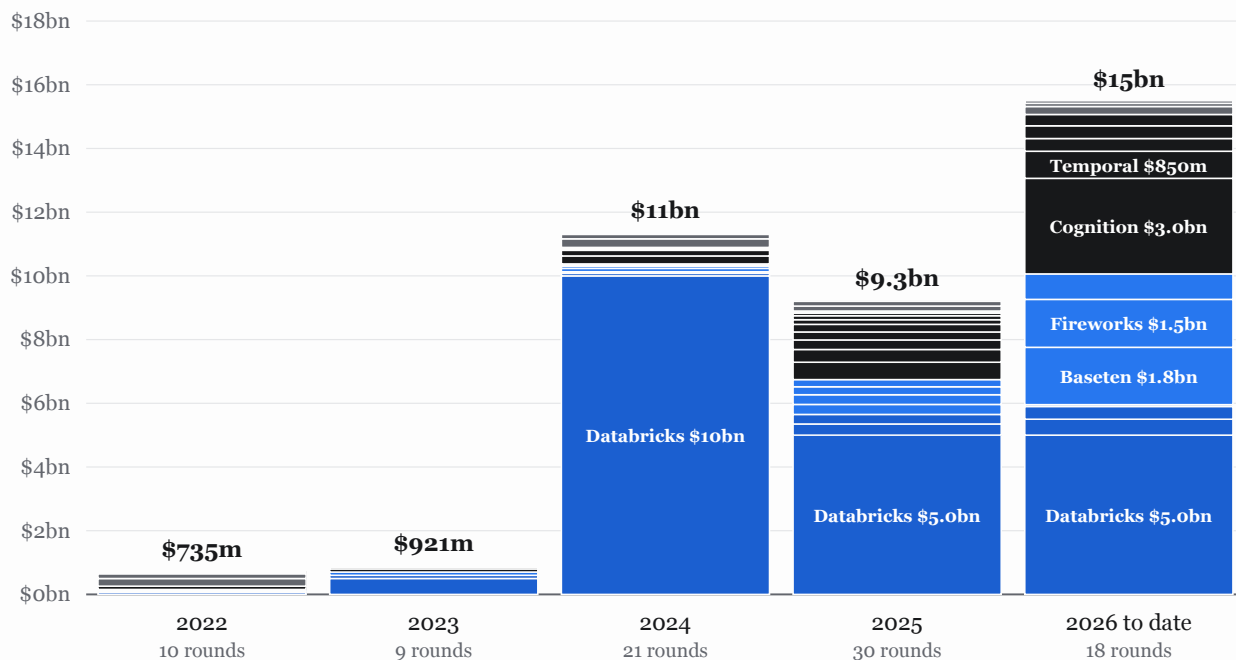
Acquisitions show the same convergence. Data platforms are adding tools to observe and control agents. Execution platforms are adding storage and identity. The boundaries in our map will move as providers compete to serve more of the work.

FIGURE 5.3

OBSERVED · DATED RECORD

Funding the companies building the backend

Capital raised by the private companies on Part Five's map, 27 of 28 with a qualifying round, each year since 2022, stacked by company



■ **State** · ■ **Inference cloud** · ■ **Runtime** · ■ **Control plane**. 88 primary equity rounds, \$38bn in all; each company's rounds in a year are one segment, and segments too thin to carry a name are the smaller rounds, all of them on the record in the data file behind this figure.

[Follow the capital raised live on the tracker →](#)

The comparison includes disclosed primary equity rounds through October 5, 2026, on the stated category definitions. Debt and secondary transactions are excluded. Financing measures capital supplied to companies, not demand for their products.

Fewer systems to connect

Building an application often means connecting separate services for data, authentication, execution and live updates. Every connection adds setup, testing and maintenance. An agent can write the code quickly and still struggle to make the whole system work.

Supabase packages services around Postgres, a widely used open-source database. Convex brings data and application functions into a managed backend. Spacetime runs application logic inside the database and streams changes to connected clients. Each reduces a different part of the work developers have to assemble themselves.

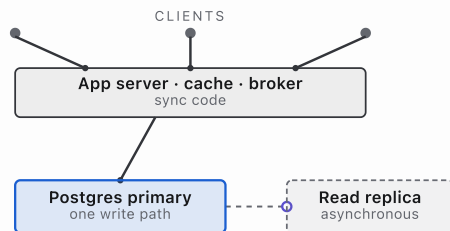
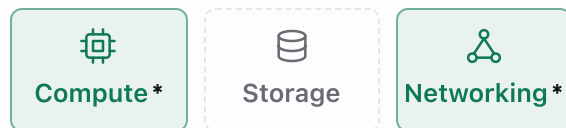
These designs put application logic and data in different places. Each reduces some work and leaves different limits. Figure 5.4 follows an update through each system, separating processing, storage and networking. A read replica is a read-only copy; WAL means the database change log. More connected users, more stored data and more writes per second are different measures of scale. The figure describes the products on the stated date, with planned features shown separately.

Different architectures move work to different places

In what ways does it scale? Each stack's write path, and the three dimensions of scale the Spacetime team names: compute, storage and networking

CONVENTIONAL STACK

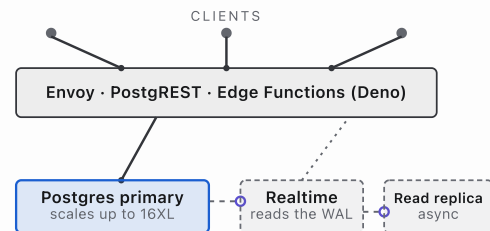
One primary takes every write; the app server, the cache and the sync code sit in front of it



* reads only, through asynchronous replicas; every write lands on one primary, whose disk is the storage.

SUPABASE

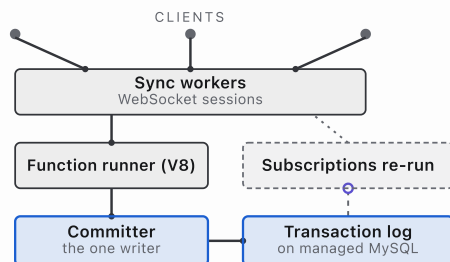
One Postgres primary; the services around it scale on their own



* writes go through one primary and replicas lag it; files scale on object storage, the database on the primary's disk; Realtime is an Elixir cluster built for millions of connections.

CONVEX

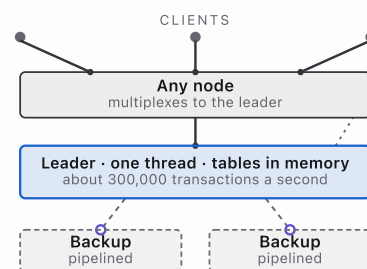
One committer serializes every write; functions that collide abort and retry



* one committer writes the log and contended transactions retry under optimistic concurrency; the sync workers hold the sessions, and how they scale is not documented.

SPACETIME

One single-threaded leader per database; replicate for durability, scale by adding databases



* one thread per database, so compute scales across databases; tables sit in memory today, with tiered storage and inter-database communication scheduled for October 31, 2026 and read replicas planned.

 scales out  scales out in part, see the line  does not ● write path

○ replicate and acknowledge ○ reads and subscriptions

Status as of October 9, 2026. Available: the core architectures drawn above. Announced: Spacetime tiered storage and asynchronous inter-database communication, scheduled for October 31, 2026. Planned: Spacetime read replicas, without a release date. Announced and planned capabilities are not available features in this comparison.

Vendor sources: [Spacetime scaling](#) · [Supabase read replicas](#) · [Supabase Realtime](#) · [Convex architecture](#) · [Convex function scaling](#). Simplified structures, not a performance ranking.

What adoption can establish

Supabase's disclosed developer base rose from about one million in October 2024 to nearly ten million in June 2026. Convex has also reported substantial growth in developers and applications. These figures show distribution and interest. They do not yet give us comparable measures of paid retention or the share of agent-built applications that work correctly.

We ask the same questions of every architecture. Can an agent build a correct application with fewer repairs? Can the system handle the resulting workload? What does it cost to operate? Do customers keep using it?

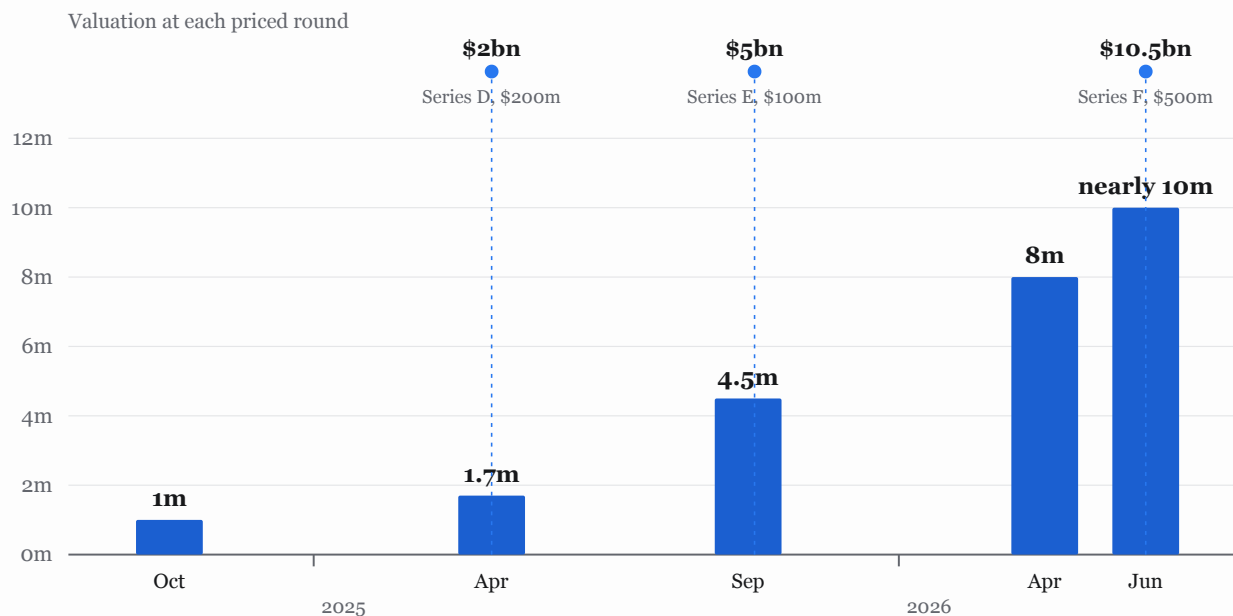
Different applications may need different answers. A large analytical workload and a small collaborative app need not run on the same design. The opportunity can support several approaches.

FIGURE 5.5

OBSERVED · REPORTED

Supabase's developer count has grown rapidly

Developers on Supabase at each disclosure, October 2024 to June 2026, with the valuation of each priced round



1m on October 3, 2024, [Fortune](#), a year before its October 2025 report · 1.7m on April 22, 2025, [TechCrunch](#), at the Series D · 4.5m on September 29, 2025, [Supabase](#), at the Lovable Cloud launch · 8m on April 2, 2026, [Supabase](#), at 100,000 GitHub stars · nearly 10m on June 4, 2026, [Supabase](#), at the Series F.

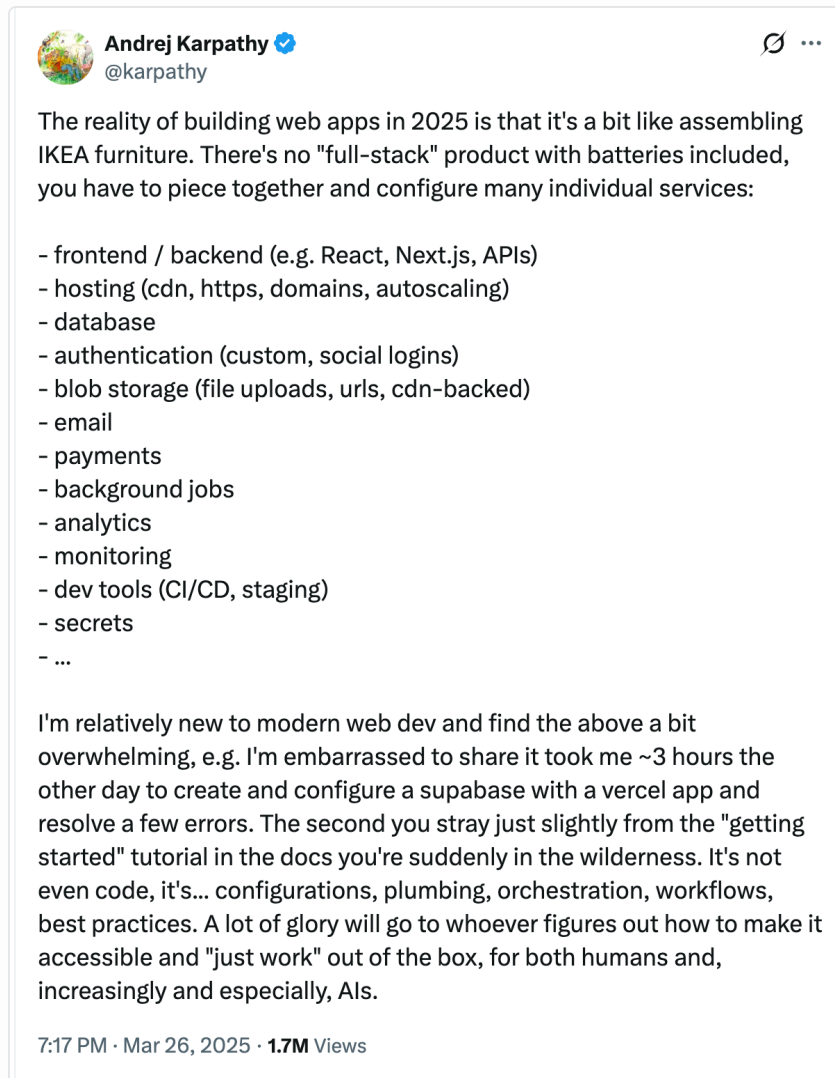
Supabase reported about one million developers in October 2024 and nearly ten million in June 2026. The chart preserves the company's definition. Registered developers are not the same as active production customers or paying accounts.

From assembling services to running software

Andrej Karpathy described the problem in March 2025: building a web application still requires a developer to assemble and configure many separate services. His post captures a constraint we have met in our own work. Generating more code does not by itself remove the work of connecting it to a reliable system.

A developer's view of the remaining complexity

Karpathy's description of the setup work between writing code and running an application



Screenshot of a personal post, reproduced for comment; copyright remains with the author.

Source: [Andrej Karpathy on X, March 26, 2025](#)

Practitioner observation. Karpathy's post describes the work of assembling web services. It gives context for the architecture question; it is not a benchmark or a measure of market-wide adoption.

Figure 5.7 shows one proposed solution. Clockwork Labs' diagram brings application logic, data and live updates together in Spacetime. The idea is the same application with fewer systems to connect—and less connection code to write and maintain.

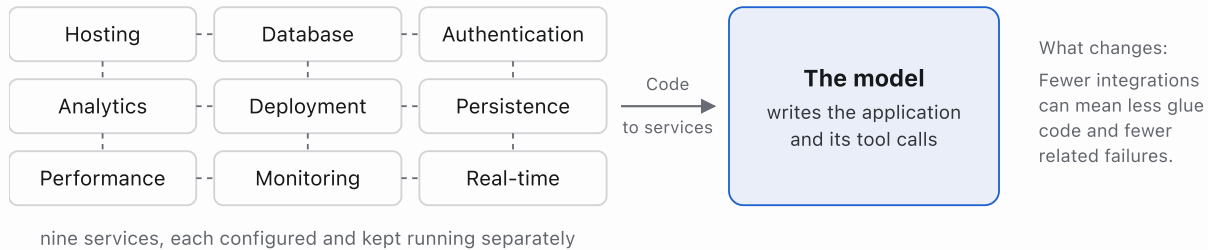
FIGURE 5.7

STRUCTURE · REDRAWN FROM THE COMPANY'S ILLUSTRATION

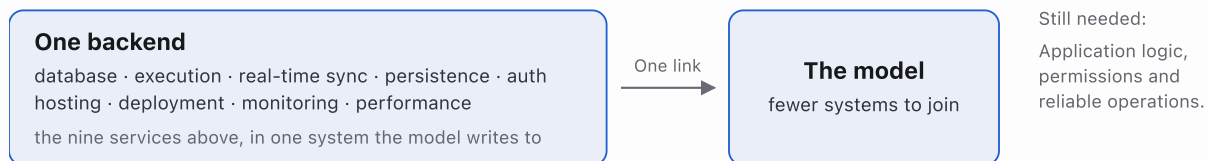
A proposed solution to the missing infrastructure piece

The setup work between a model and a running application, before and after one backend takes it on; the company's framing, with the services it names

BEFORE · ASSEMBLED BY HAND



AFTER · ONE BACKEND, THE COMPANY'S PROPOSAL



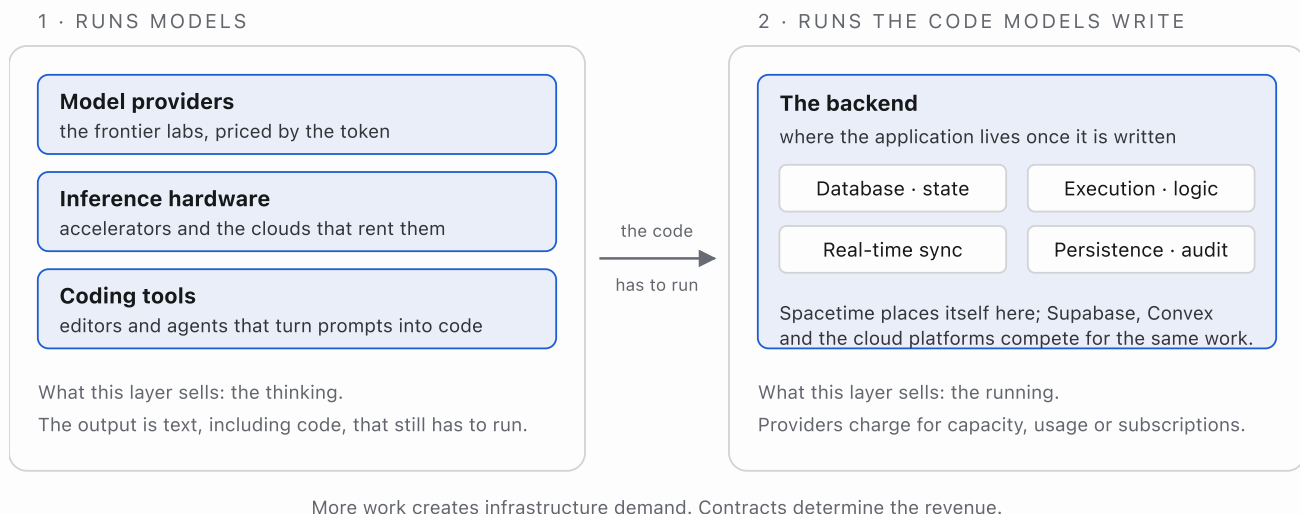
The next diagram separates two jobs. One layer runs the models. Another runs the code they write. Both are needed when an answer becomes an application that people depend on. The second layer has to keep working after the conversation ends.

FIGURE 5.8

STRUCTURE · REDRAWN FROM THE COMPANY'S SLIDE

Running models and running the code they write

The systems that run models and the systems that run the code models write, with categories in place of the slide's logos



What we have learned by building

Machine Hour runs on Spacetime. [Skycatcher](#) is an investor in Clockwork Labs, the company behind it. That relationship informs our access and our interest; it also makes clear disclosure and independent testing important.

Using the product gave us a direct view of the architectural idea. Application logic and data sit together, and connected clients receive changes through subscriptions. An agent has fewer separate services to configure. Our experience shows that this approach can serve this site. Larger and more demanding applications need their own evidence.

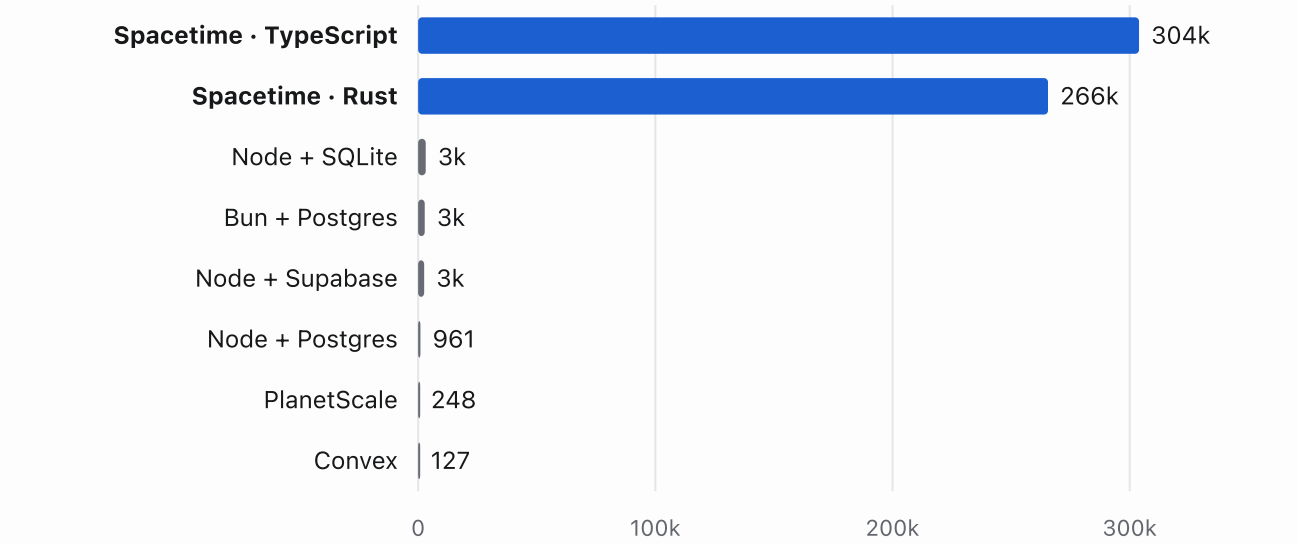
Clockwork Labs has published tests of transaction throughput and application building. The first measures a specific workload. The second asks the same agent to build a twelve-feature chat application on two backends. The figures below show the results and the limits of each test.

Independent replication, comparable configurations and paid applications running reliably over time would strengthen this evidence.

FIGURE 5.9

A vendor benchmark on one workload

Transactions a second, the colocated backends on one \$1,000-a-month node and PlanetScale at its own cloud price, as Clockwork Labs measured it on May 14, 2026



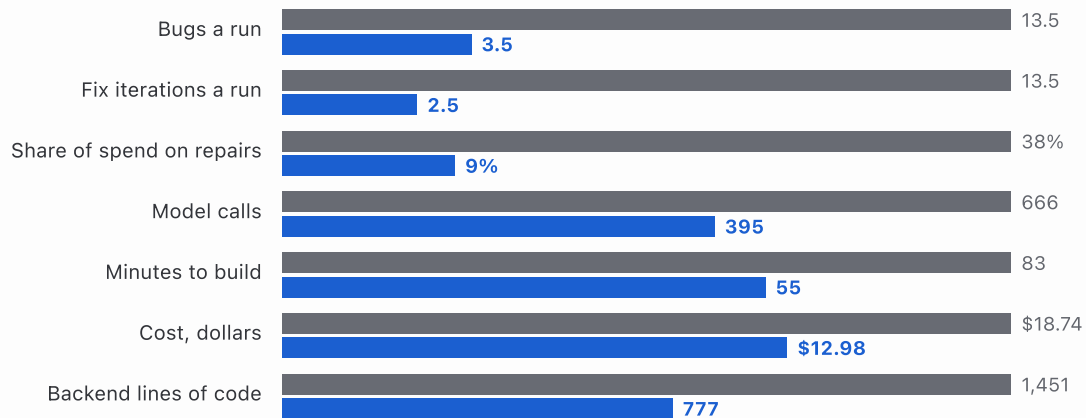
Transactions a second, linear scale, on the vendor’s contended fund-transfer workload. The numbers are Clockwork Labs’ own, from [its May 14, 2026 post](#); they are not independently verified. [Skycatcher](#) is an investor in Clockwork Labs.

Vendor benchmark. Clockwork Labs reports roughly 304,000 transactions per second for Spacetime Standalone; its selected alternatives range from about 130 to 3,200. Hardware, workload and durability settings affect the result. The test is not independently verified and does not establish a general performance advantage. [Skycatcher](#) is an investor in Clockwork Labs. Method and source: [Clockwork Labs benchmark post](#).

A small test of building with agents

One model built the same twelve-feature realtime chat application on Spacetime and on a Postgres stack, one feature at a time; what it cost the agent, and what it shipped

What it took the agent · lower is better



What the agent shipped · higher is better



■ Postgres with Express, Socket.io and Drizzle ■ Spacetime

Claude Sonnet 4.6, twelve features one at a time, two runs each, averaged; the vendor's runs and grading.

Vendor benchmark. Two runs per backend, using Claude Sonnet 4.6 and manual grading against twelve requested features. This sample cannot establish a stable success rate or general architecture advantage. Disclosed prompts, repeated trials and independent grading are needed. [Skycatcher](#) is an investor in Clockwork Labs. Method and source: [Clockwork Labs benchmark page](#).

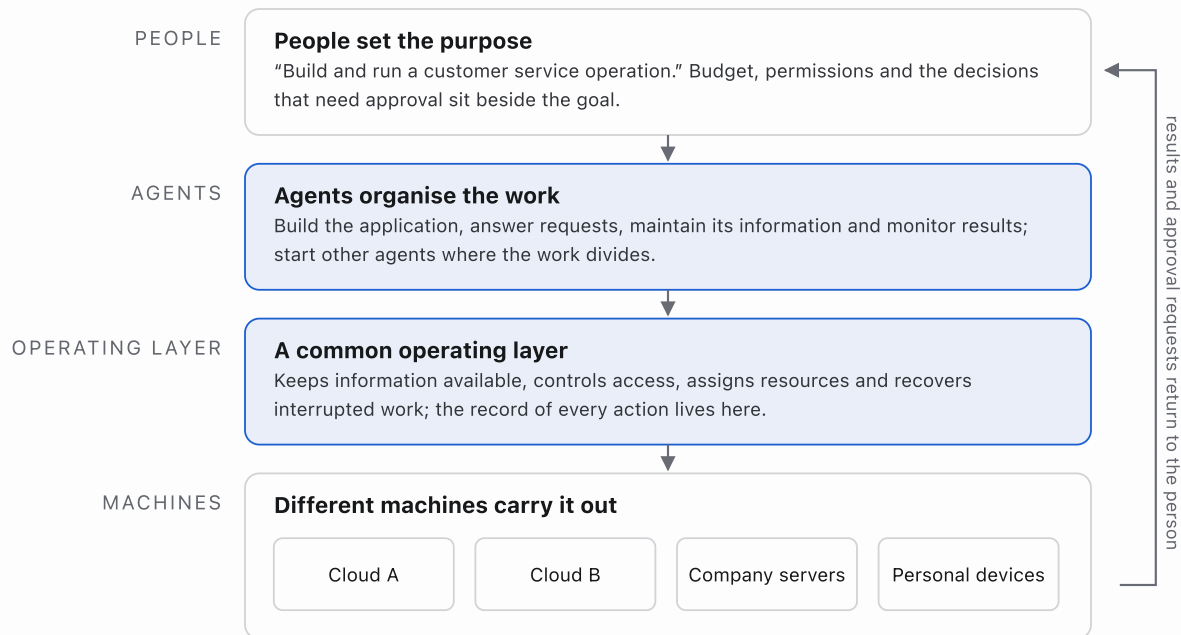
An operating layer for the machine economy

Imagine describing a service you want to exist, and having agents build it, launch it and keep it running across clouds and devices. You set the goal, budget and decisions that need approval. A shared operating layer preserves the data, permissions and history so another agent can carry on when one stops.

That layer needs four capabilities: place work on suitable, authorized hardware; preserve information between tasks; recover from failures; and enforce people's permissions, budgets and approval rules. It could connect several platforms through shared standards.

One goal, many machines

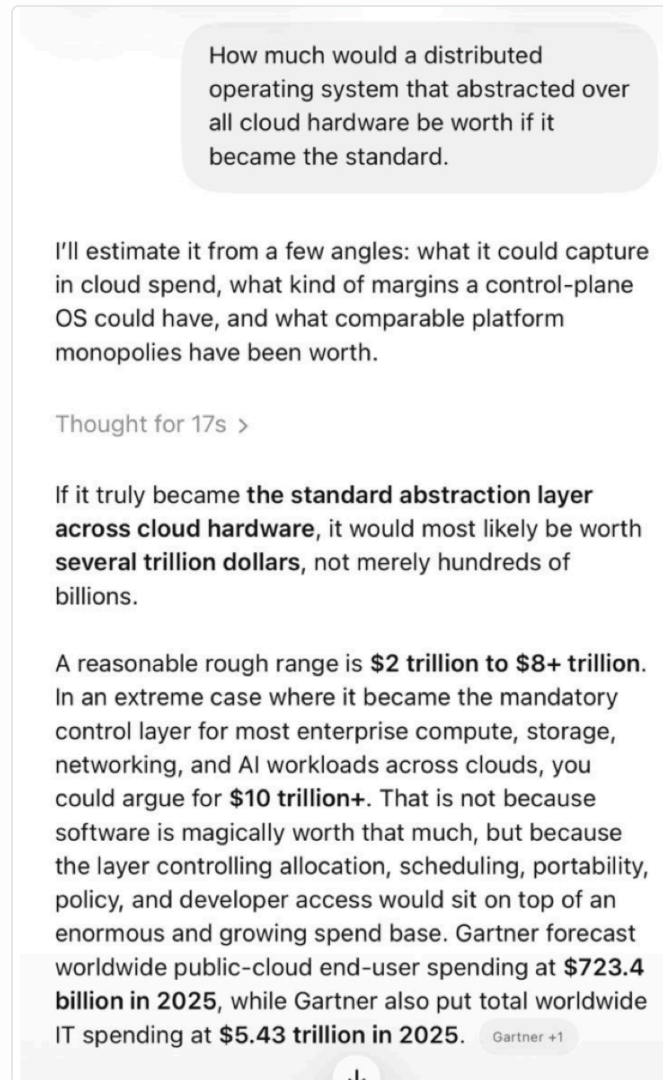
A possible architecture: agents coordinate work across authorized computing resources, while people set goals and limits



Illustrative vision, not a product or a measurement. A possible architecture: agents coordinate work across authorized computing resources, while people set goals and limits, and results and approval requests return to the person.

In April 2026, we asked ChatGPT to imagine the scale of such a platform. The exchange below is a thought experiment. Its useful question is how much computing could pass through a common operating layer.

The blue-sky scenario: a common operating layer



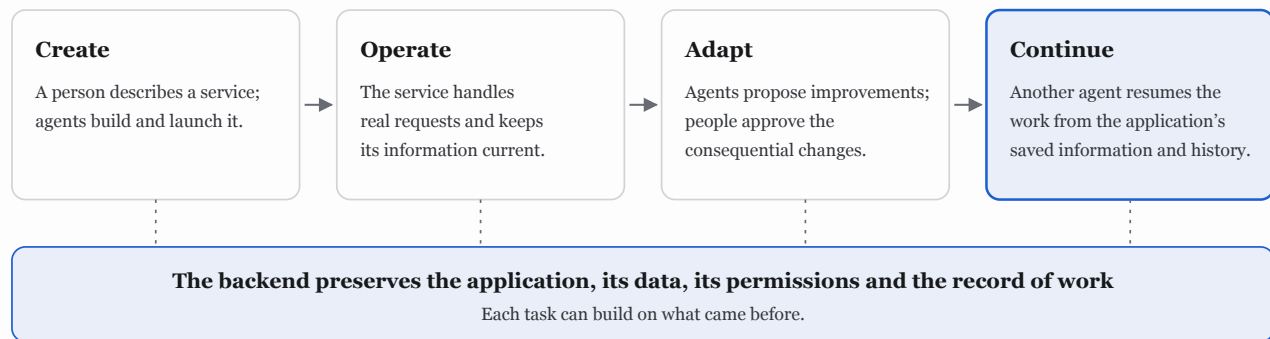
A ChatGPT answer to the question shown, April 2026, reproduced for comment. A model's estimate, not ours and not a forecast.

Thought experiment. This April 2026 ChatGPT exchange considers a distributed system across cloud hardware. Its valuation figures are speculative model output, not an independent valuation, a price target or evidence of demand.

The lasting value would come from useful systems that keep operating. Agents build and launch a service, maintain its information and propose improvements. People approve consequential decisions. If a task is interrupted, another agent can resume from the saved data and history.

The work continues after the conversation ends

One application through create, operate, adapt and continue, on a backend that keeps it



Illustrative vision. One application through four stages, create, operate, adapt and continue, on a backend that preserves the application, its data, its permissions and the record of work. The lasting value comes from useful systems that keep operating; each task can build on what came before.

What would make that vision real

The hard part is making these systems work together reliably: moving data, isolating customers, recovering from failures and controlling cost and permissions across machines. Portability must work in production. No single company has established this outcome.

Our machine-hours thesis does not require this full vision to arrive. It describes what sustained growth in machine work could make possible.

We believe people will increasingly direct work that agents carry out and maintain. The backend would become the foundation for that work: keeping applications running, preserving what has been learned and allowing the next task to build on the last. The larger possibility is a world where far more people can turn an idea into something that works, and keeps working.

Evidence, methods and prediction records

The five parts make the argument. These appendices provide the definitions, evidence and assumptions behind it. They include token records, the hours model, company filings and the original prediction terms. The final group summarizes company presentations in our own illustrations, with links to the original sources. Use the links below to go directly to a source or method.

Glossary

A · Tokens: the history, our forecast and the outside view A1 Tokens, from the beginning

A2 Our token forecast: about 2,300 quadrillion a year by 2029

A3 The outside forecast: Situational Awareness, graded

A4 Estimating current machine hours and the forecast starting point

A5 The human benchmark: 7.0 trillion paid hours a year

B · The machine hours model: the assumption register and Jevons' paradox

B1 The assumption register

B2 Jevons' paradox: the illustration, the precedents and the elasticity condition

B3 Displacement: machine work replacing human work where both can be counted

B4 The proxies: the dated disclosures behind the inputs

C · The companies: the roster, the value chain and the private record C1 The full research roster

C2 From the value of the work to backend revenue C3 The private universe: the dated record

C4 State and Work: the economic questions

D · The record: downside, registered calls and disclosures

D1 What would change our view: sensitivities, risks and evidence to watch

D2 The launch predictions and their grading rules D3 Disclosures

E · Company presentations: findings summarized in our own illustrations

E1 Cloudflare, June 9, 2026 E2 Datadog, February 12, 2026 E3 MongoDB, September 29, 2026

E4 Snowflake, June 2, 2026

Glossary — plain words

TERM	MEANING
Machine labor	Work that AI agents plan, execute and check toward a goal.
Machine hour	One hour an agent spends actually working, with waiting and sleep left out and parallel workers counted separately. The unit the model runs in.
State	Data and memory: databases, storage and search. Five public companies in the universe.
Work	Execution and control: runtimes, networks, monitoring, identity and security. 27 public companies.
Applications	The 49 companies that deliver outcomes above the backend. Mapped for context; not in the universe.
The stack	The five layers: Applications, Work, State, Intelligence and Computation, with the hyperscalers and SpaceX mapped beside them as a cross-layer group. 108 public pure plays.
The backend	State and Work together: the 32 public companies the universe, the indices, the medians and the bets run on.
Intelligence	The layer that sells tokens and inference. Two listed pure plays, Z.ai and MiniMax, both in Hong Kong. Mapped for context; not in the universe.
Computation	Silicon, memory, networking, systems, power and rented capacity. 25 public companies. Mapped for context; not in the universe.
The four company cases	Cloudflare, MongoDB, Snowflake and Datadog, one for each activity of an agent hour: it executes, remembers, asks questions of data and answers for itself.
Provider use	The assumed share of machine hours using a provider. One hour can use several providers, so their shares overlap.
Illustrative bill per hour	The list-price cost of the resources used in one modeled hour, before discounts. Actual bills depend on contracts and use.
Price at ten times the volume	The model’s price path: what a unit costs, relative to today, when volume is ten times higher. Applied per tenfold, so at a hundredfold it falls twice.

TERM	MEANING
Existing-business revenue	Revenue outside the separately modeled agent workload, including existing AI use. The Agentic Workforce holds reported growth through 2028; the Machine Economy adds three percentage points a year from 2027 through 2029. Both then fade to their stated 2032 rates.
Agent share of revenue	Agent revenue divided by total revenue: under 1% in 2025, about 2% to 3% by 2029, and about a fifth to two fifths by 2032 in the two scenarios.
The two scenarios	The Agentic Workforce, where most agents stay human-directed, and the Machine Economy, where agents start work on their own. Both reach a machine majority of hours.
Human-directed agents	Agents a person starts and steers. One per user at about 12 minutes a day in 2026, rising to 16 at 1.8 hours in 2032, the same in both scenarios.
Autonomous agents	Agents that start work on their own. In 2032, 5 per user in the Agentic Workforce and 45 in the Machine Economy; this is where the two scenarios part.
Always on	An autonomous agent that runs most of the day: 17 hours in the Agentic Workforce and 23 in the Machine Economy in 2032, against 3.5 and 5.8 hours for the rest. 25% of autonomous agents are always on in the Agentic Workforce, 70% in the Machine Economy.
The consensus case	Analysts' consensus revenue for the four to 2029, then growth fading to 12% a year by 2032. The bear case fades growth to 3% by 2029.
Deferred revenue	What customers have been billed or have paid ahead of use; the contract liability as reported. It moves before recognized revenue.
RPO	Remaining performance obligations: all contracted revenue not yet recognized, broader than deferred revenue.
Net revenue retention	What last year's customers spend this year as a share of what they spent then, on the vendor's own definition.
Free cash flow	Operating cash flow less capital purchases. Share dilution is counted separately.
Bold bet	A dated call with a measure and a deadline. We grade it once at its deadline and publish the grade beside it.

Appendix A — Tokens: the history, our forecast and the outside view

A1 · Tokens, from the beginning

A token is a small piece of text a model reads or writes. We use token disclosures because they provide a public record of model activity. They are not direct measurements of working time or useful output.

The historical series brings together disclosures and estimates with different coverage. Comparisons must use the same token basis and identify changes in coverage. Fresh input, cached context and written output should be shown separately where the source allows it.

The broad record shows rapidly rising use alongside falling prices. A 2,000-fold volume increase and a 200-fold price decline produce a descriptive log ratio of about 1.4. They do not identify price elasticity: better models, adoption and changing workloads also affected demand.

The sensitivity illustration asks what happens at different assumed responses to price. It does not claim that one response has been measured for the whole market.

PRICE PER ACCEPTED TASK	DEMAND AT ELASTICITY 0.7	DEMAND AT 1.0	DEMAND AT 1.3
1.00×	1.00×	1.00×	1.00×
0.50×	1.62×	2.00×	2.46×
0.10×	5.01×	10.00×	19.95×

A2 · Our token forecast: about 2,300 quadrillion a year by 2029

The token forecast checks the hours scenario against another unit. Multiply machine hours by the assumed tokens per hour. In the Machine Economy, about 2.9 trillion hours in 2029 at roughly 800,000 fresh tokens per hour gives approximately 2,300 quadrillion tokens a year: about 130x the 2026 base of about 17 quadrillion, or about 5.1x a year. The fall from about 1.4 million fresh tokens an hour in 2026 to 800,000 assumes that routine steps move to smaller models and that more context is served from cache as tools mature; it is an assumption, not a measured trend, and the recorded monthly series will show whether it holds. A quadrillion is one thousand trillion.

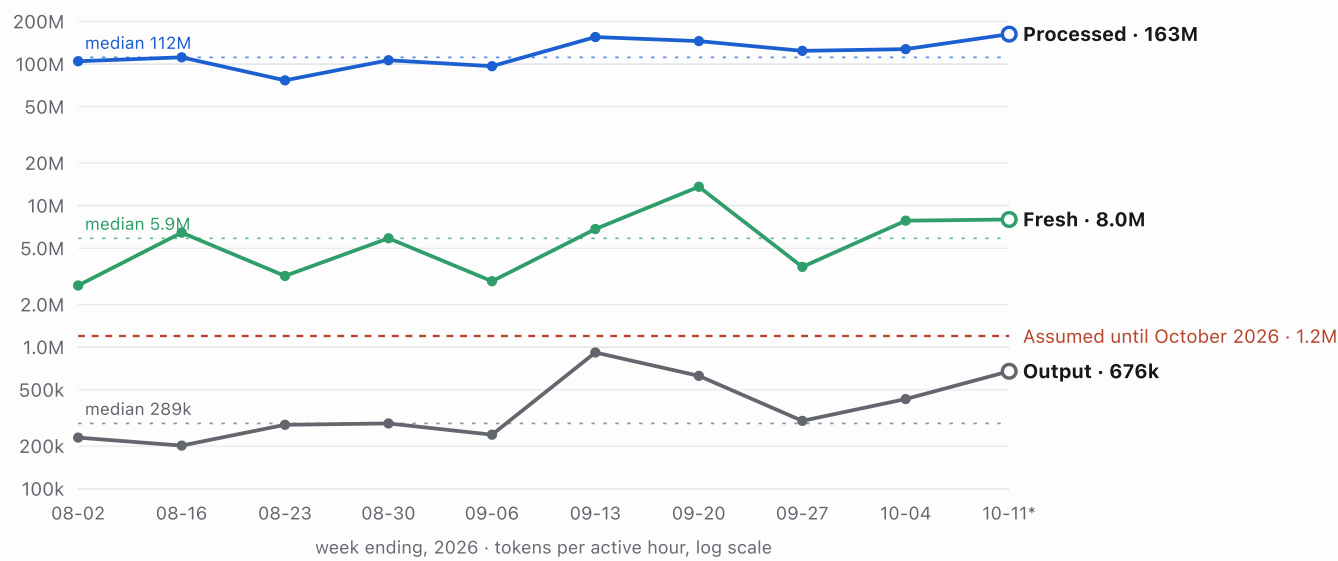
A second calculation starts with users. Roughly two billion weekly users in 2029, 22% running continuously active agents and 10–25 million tokens per participating user per day produce a broad range around the hours-based result. These are assumptions, so agreement between the calculations is not independent confirmation.

Our coding logs provide calibration evidence. Since October 7, the published measure counts intervals when the model generates or a tool runs. It excludes waiting for a person and counts a parent's wait for a child only in the child's execution. The earlier capped-gap estimate remains in the record so the effect of the change is visible.

The model converts at about 1.4 million per machine hour in 2026, the blended fresh-input rate across coding, research and enterprise workflows. It updates with the data and is recorded month by month. The original 2.3-million figure describes the coding sample used to set the hours anchor on October 6; it remains in the record. Fresh input, cached reads and output are shown separately so a definition change cannot look like an improvement in efficiency.

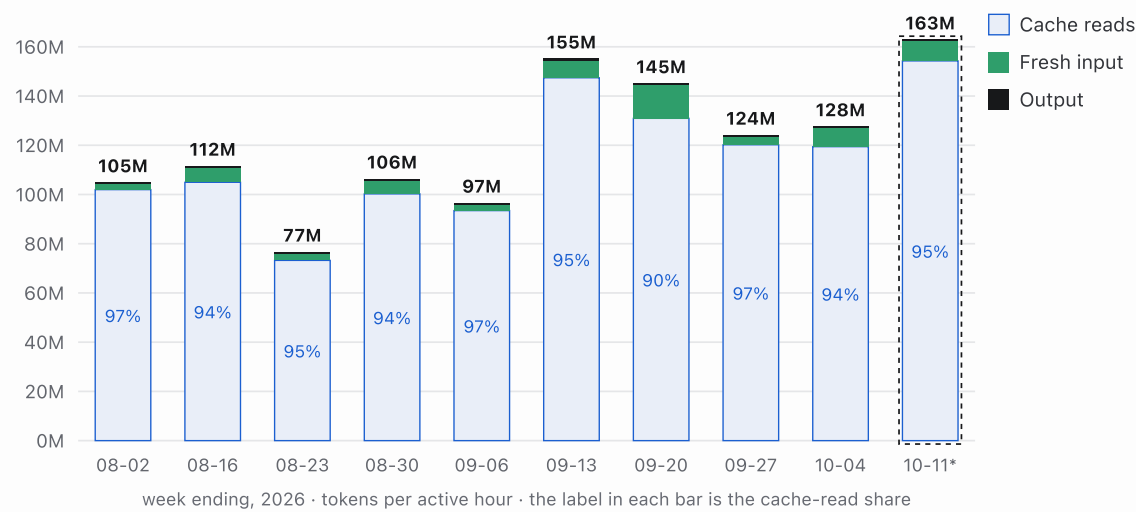
Over the nine complete weeks through October 4, 2026, the published series shows about 257 calls an hour, a median 112 million processed tokens and a median 5.9 million fresh input tokens per active hour. Cached reads account for 95% of processed tokens. On the archived capped-gap estimate the same weeks held about 1.8 times as many hours, so every rate read lower (about 144 calls, 61 million processed and 3.4 million fresh an hour); the restatement is method, not growth. The one-operator processed divisor the gauge once used for OpenRouter's traffic, the measured cadence times the platform's 100,000 tokens a request plus output, reads about 28 million on this basis. These observations describe our coding sample. The hours anchor remains subject to the A4 review rule.

Tokens per active agent hour, week by week: processed, fresh and output, with the constant the model assumed before October 2026



Processed tokens include everything read and written. Fresh input includes new input and cache writes; output is what the model writes. The model converts at 1.4M fresh tokens per machine hour in 2026, blended across workloads and updated with the data. The original 2.3-million figure describes the coding sample used to set the hours anchor on October 6. The chart shows the coding operator’s logs on the execution-interval method introduced October 7: time when a model generates or a tool runs, excluding waits for a person. It is distinct from the older capped-gap estimate, which the measurement file keeps beside it. Weeks end Sunday. An asterisk marks a partial week; medians use complete weeks only.

Tokens used per active hour, by kind: cache reads, fresh input and output



The height of each bar is what a provider such as OpenRouter counts; fresh input and output together are the thin band at the top, three to eight percent of the hour. Measured. An asterisk marks a partial week; its bar is outlined with a dashed border.

2B weekly users

Assumed addressable weekly AI-user population in 2029.

22% daily-agent participation

440M participating users; this is a behavioral assumption.

10–25M tokens per active day

One workload per participant × 365 active days/year.

1,606–4,015Q

annual tokens · separate illustration

A3 · The outside forecast: Situational Awareness, graded

We compare the hours forecast with an illustrative supply path and with the themes in Leopold Aschenbrenner’s Situational Awareness. His published predictions and our calculations using related assumptions are labeled separately.

An announced campus is not operating inference capacity. The evidence record distinguishes announced investment, installed hardware, available power and capacity serving workloads.

The conversion from tokens to hours depends on workload mix and the compute required for fresh input, cached context and output. We show a simplifying arithmetic case and a price-weighted stress case. Prices are not physical compute ratios, and cached reads are not physically free.

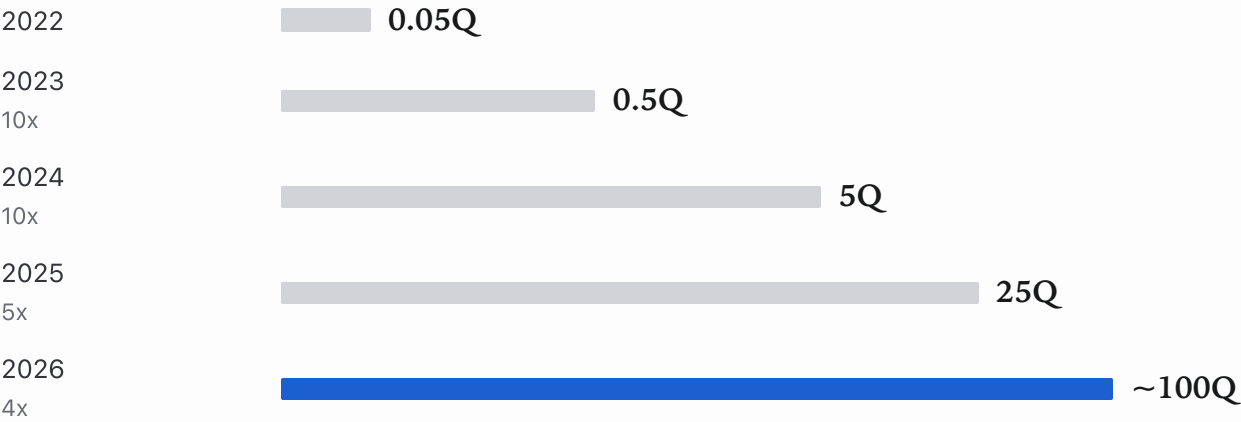
The published 2029 path serves roughly four trillion machine hours annually, or nearer three trillion on the stress test. From 2029 to 2032, the capacity path assumes eight times as much compute and tenfold greater token efficiency. The stricter case also assumes less repeated context per fresh token. The charts show requested and delivered work, so a supply constraint also reduces the related revenue estimate.

A separate illustrative calculation uses Epoch’s estimates of installed AI chips and DeepSeek’s disclosed throughput. It asks whether the order of magnitude is plausible. The result depends on inference’s share of capacity, utilization and the models served. It is not a measured ceiling or proof that future power and hardware will be available.

Tokens a year: reported and estimated history, then the three places 2029 can land

Reported and estimated tokens a year, 2022 to 2026

QUADRILLIONS OF TOKENS A YEAR (Q), BARS ON A LOG SCALE; THE MULTIPLE IS THE RISE ON THE YEAR BEFORE



Forecast: tokens a year in 2029, three destinations from the same base

QUADRILLIONS OF TOKENS A YEAR (Q)

2026 base

processed tokens as disclosed; the forecasts count fresh tokens, the unit the hours model converts

~100Q

Skycatcher forecast

about 130x the model's 2026 base of about 17Q fresh tokens; the user-based sensitivity in A2 spans 1,606 to 4,015Q

2,300Q

User-based

sensitivity

our inputs, not Aschenbrenner's figures: 2 billion weekly users, 30% participation, 25 to 30 million tokens a day · 5,475 to 6,570Q

~6,000Q

Processed-token

ceiling

if the announced build lands: 10x power × 3x software × 2.5x per watt; the fourth exhibit converts it to compute

~7,500Q

MILESTONE	EVIDENCE NEEDED	
Announced capital	Dated announcement and defined scope	
Delivered equipment	Installed and commissioned capacity	
Energized power	Operating power at the relevant facility	
Inference delivered	Measured serving capacity and workload mix	
ASSUMED INPUT	LOWER	UPPER
Weekly AI users	2 billion	2 billion
Daily-agent participation	30%	30%
Tokens / participating user-day	25 million	30 million
Annual tokens, 365 active days	5,475Q	6,570Q

Fresh tokens to compute: the forecast and the announced build in one unit, a calculation and a stress test, then capacity after 2029

2032 demand / available capacity: 340.0 / 357.4 trillion hours a year. The Machine Economy uses about 95% of this illustrative allowance. It requires compute to double each year after 2029, tokens per hour to fall tenfold, and nearly the full modeled allowance to be available to agents.

Step one, the workload: one agent hour on our trace reads 5.88M fresh input tokens, re-reads 106.04M cached tokens and writes 0.29M (complete-week medians, one operator, subagents included). Step two, the mix of the world’s 2026 base of about 100Q processed tokens: assumed 40% fresh input, 50% cached re-reads, 10% written, since no provider publishes it. Step three, compute per token on two weightings. Step four, both sides in compute units: the ceiling of ~7,500Q processed tokens, the forecasts of 2,300Q and 470Q fresh tokens, and the Machine Economy’s 2032 demand before and after the assumed efficiency gain. Both comparisons use the same 2029 computing ceiling; the separate table below adds later computing growth. The calculation is the arithmetic column; every surface of the site uses its cap.

STEP	THE CALCULATION: BY ARITHMETIC A CACHED TOKEN COSTS NOTHING TO RE-READ, A WRITTEN TOKEN COSTS ONE	A STRESS TEST: BY LIST PRICE CACHED A TENTH, WRITTEN FIVE, THE RATIOS ON ANTHROPIC’S PRICE LIST; TOKEN PRICES ARE NOT COMPUTE RATIOS
Compute per fresh token, agent profile	1.05	3.05
Compute per processed token, world base	0.50	0.95
2029 ceiling, in compute	3,750Q	7,125Q

STEP

THE CALCULATION: BY ARITHMETIC
A CACHED TOKEN COSTS
NOTHING TO RE-READ, A
WRITTEN TOKEN COSTS ONE

A STRESS TEST: BY LIST PRICE
CACHED A TENTH, WRITTEN FIVE, THE
RATIOS ON ANTHROPIC'S PRICE LIST;
TOKEN PRICES ARE NOT COMPUTE
RATIOS

Machine Economy 2029 · share of
ceiling**2,413Q · 64%****7,015Q · 98%**Agentic Workforce 2029 · share of
ceiling**493Q · 13%****1,433Q · 20%**

Hours the 2029 build serves, a year

4.5 trn**2.9 trn**2032 demand at 2029 token intensity
÷ 2029 computing ceiling**76x****116x**2032 demand after assumed efficiency
gain ÷ 2029 computing ceiling**7.6x****11.6x**

At 2029 token intensity, 340 trillion machine hours would require 271,979Q fresh tokens. The 2032 scenario assumes ten times fewer tokens per hour, reducing that requirement to 27,198Q. Its capacity case depends on that efficiency gain as well as continued computing growth. One Q is one quadrillion tokens.

CAPACITY-PATH
ASSUMPTION

2029

2030

2031

2032

Fresh tokens per
machine hour**800,000****372,093****173,913****80,000**

Illustrative capacity conversion, not a physical measurement: cached context still takes memory and transfer, and the arithmetic column prices it at zero; calibration against measured throughput by model, context length and utilization is the work ahead. The tenfold fall in tokens an hour means about 80,000 fresh tokens a machine hour in 2032, from 800,000 in 2029, the basis the token forecast keeps for 2029. The 2029 build alone, about 4.5 trillion hours a year, sits below the 7.0 trillion crossing, so the prediction itself needs the build to keep growing: one more doubling crosses in 2030. The Agentic Workforce needs the doubling to continue; the Machine Economy needs the doubling and the tenfold fall in tokens an hour together, and on the stress-test weighting also the context re-read per fresh token falling by about half. The published case assumes the whole of that capacity is available to agents; the model page carries the share as a control, and revenue there follows the hours delivered. At half, the Machine Economy's 340 trillion exceeds what the build serves in 2032 and the case delivers about 179 trillion; the 7.0 trillion crossing survives at any share above about 2% of the 2032 capacity, though a smaller share delays the year. Moving the world's cached share between 30% and 70% moves the ceiling by about a fifth on the calculation.

CAPACITY AFTER 2029, HOURS A YEAR THE BUILD SERVES (CALCULATION
COLUMN)

2030

2031

2032

The 2029 build, held flat

4.5 trn**4.5 trn****4.5 trn**

The build doubling each year

8.9 trn**17.9 trn****35.7 trn**Doubling, and tokens an hour falling tenfold by 2032 (phased): the
model's capacity path**19.2 trn****82.2 trn****357 trn**

The same, with half of it available to agents

9.6 trn**41.1 trn****179 trn**

The same, with a quarter available to agents

4.8 trn**20.6 trn****89.4 trn**Demanded, from the model: the 7.0 trillion crossing · the Agentic Workforce ·
the Machine Economy**7.0 · 2.0 ·
15.8****7.0 · 6.3 ·
77.4****7.0 · 19.0 ·
340**

The capacity path multiplies the 2029 ceiling by 8 (the build doubling each year to 2032) and by 10 (tokens a machine hour falling tenfold), 80x in all. The share of 2032 capacity the crossing needs on the model page is a share of that figure, so it assumes both conditions.

INDEPENDENT CHECK FROM THE HARDWARE RECORD	VALUE	SOURCE
AI compute deployed, end of 2025	12 to 16 million H100-equivalents; 14 million used	Epoch AI, May 20, 2026
Growth of the stock to 2029	2x a year (the model's doubling) · 3.3x a year (Epoch's production trend, a doubling every seven months)	Epoch AI, January 9, 2026
Stock at the end of 2029	224 million · 1,660 million H100-equivalents	calculated
Throughput of one GPU	9,213 input tokens a second on prefill, 1,850 written tokens a second on decode (one H800 node of 8, February 2025)	DeepSeek, March 1, 2025
Share serving inference · utilization	50% · 50%	assumed
Compute units one GPU serves a year	81 billion fresh and written tokens	calculated, on the world mix of fresh to written
The stock serves in 2029	9,059Q · 67,145Q	calculated
The price-derived ceiling, same units	3,750Q	the calculation column above
Hardware check ÷ ceiling	2.4x · 18x	calculated
Machine hours that stock serves in 2029, at the model's 2029 tokens an hour	11 · 80 trillion a year, against 4.5 on the ceiling	calculated

A plausibility check, not a second ceiling: the stock and trend are Epoch AI's ([deployed compute, May 20, 2026](#); [production trend, January 9, 2026](#)), the throughput is one disclosed system ([DeepSeek V3/R1 serving, March 1, 2025](#)) on 2025 hardware, and the inference share and utilization are assumptions. It says the price-derived ceiling is a conservative reading of the announced build: on the model's own doubling the hardware serves more than the ceiling counts, and on Epoch's trend several times more. What binds the Machine Economy after 2029 is the growth of the build and the fall in tokens an hour, not the stock that exists.

A4 · Estimating current machine hours and the forecast starting point

We use two main estimates and a separate cross-check.

The **provider-by-provider estimate** estimates hours venue by venue: token volume × estimated agent share ÷ tokens processed per agent hour. We account for routed traffic to avoid counting it at both OpenRouter and the underlying model provider. The October 8, 2026 register defines the venue divisors as assumptions about active model and tool execution, excluding unattended idle time and between-turn human waiting. They are informed by workload profiles, not a matched timing sample across each venue. They should not be read as measured venue-wide rates or as the older capped-gap timing estimates. The original forecast anchor remains fixed under the rule below.

The **OpenRouter-based estimate** converts classified agent tokens into hours by workload, then scales those hours by OpenRouter's share of worldwide tokens. It assumes the platform's agent mix helps describe the wider market. That assumption is especially important because OpenRouter is

not a representative global sample.

The **billed-activity cross-check** converts selected product revenue or work units into illustrative hours. Cognition's revenue includes products beyond time-based agent work, and the duration assigned to a Salesforce work unit is our assumption. The result is not a measured lower bound.

For the week ending October 18, 2026, the latest OpenRouter-based estimate is 55 million machine hours a day. The alternative provider-by-provider estimate is 17 million. Neither is a direct count. Assigning all uncertain task tags to coding gives 45 million, to enterprise workflows 55 million and to research 88 million. Varying workload-profile endpoints gives 25 to 154 million; this range varies tokens used per active hour only. It does not capture all uncertainty in worldwide token volume, agent classification or how well OpenRouter represents other platforms. It is not a confidence interval.

The workload mix is dated October 17, 2026. Its central fresh-input rate is about 1.3 million an hour across the mix. Applying the coding operator's fresh-token rate alone would imply about 29 million hours a day. That comparison shows why a single coding trace should not stand in for all agent work.

Were the forecast restarted from the latest OpenRouter-based estimate, 55 million a day for the week to October 18, 2026 instead of the published 35 million, the Agentic Workforce destination would rise from 19 to 30 trillion hours and its crossing would move from 2032 to 2031, with delivered hours held at the capacity path. This is an illustration. A change to the forecast must follow the review rule below.

These methods share evidence and assumptions. Their range is not a confidence interval. Work outside the covered venues and tool execution with little token use remain difficult to observe.

The published 2026 anchor remains 35 million hours a day. At a quarterly review the two estimates are compared with the forecast for the same date, not with the fixed 2026 anchor, since ordinary growth along the published path is expected to carry the readings above 35 million. If both estimates have sat on the same side of the forecast for a full quarter, the published growth rates are held and the forecast is restarted from that quarter's pulse as a new vintage, dated and recorded; if the gap is instead traced to a change in how 2026 was measured, the 2026 anchor is restated and the path redrawn from it. Either way the registered prediction keeps its original terms.

A revised method restates history on a comparable basis and preserves the earlier series. We record the difference as a measurement change rather than growth.

From OpenRouter's hours to a world figure, step by step, with the model path for scale

METHOD	ARITHMETIC	MILLION HOURS A DAY
Provider-by-provider estimate	12 venues: tokens a day × agent share ÷ tokens an agent hour, summed; OpenRouter's estimated worldwide share and hours calculated by workload, counted once (its 4.1% of the world's tokens taken out of the pooled venues), the rest graded (B1)	17
Inference clouds (Fireworks, Together, Microsoft Foundry)	52.5T a day × 60% ÷ 7.5M an hour	4.2
Anthropic (estimate)	157.2T a day × 65% ÷ 30.0M an hour	3.4
ByteDance (Doubao)	172.6T a day × 10% ÷ 6.0M an hour	2.9
Google	100.8T a day × 15% ÷ 6.0M an hour	2.5
OpenRouter	23.7T a day × 73% measured ÷ 7.5M an hour calibrated by workload	2.3
OpenRouter-based estimate	17.0T agent tokens a day split by tag into coding 49%, research 15%, workflow 35% (uncertain tags allocated, next row), each ÷ its profile's tokens an hour (B1), summed (equivalent divisor 7.5M an hour) ÷ 4.1% share; 25 to 154 at the profiles' dense and thin ends	55
Allocation of the uncertain tags	29% of the platform's agent tokens carry tags that do not say which workflow they belong to (planning, tool dispatch, memory, classification, data extraction); allocated in proportion to the classified shares. All to coding: 45; all to workflow: 55	
Fresh tokens the hours imply	derived from the hours, not a second estimate of them: 1.3M fresh an hour across the mix, 72.2T a day; the one-operator fresh basis the anchor was set on (A2) read 29	
Activity converted to hours, a cross-check	Devin (Cognition): annualized revenue run rate passes \$1B; Salesforce Agentforce: 3.2 billion work units in Q2 FY27 (quarter ended July 31, 2026), up 97% quarter over quarter	0.89
Original 2026 forecast starting point	people × agents × hours: the published scenario anchor, set on October 6, 2026 on the one-operator fresh basis and kept until the calibration above is reviewed under the rule in A4	35

Read at the week to 2026-10-18. The two estimates address one quantity. The OpenRouter-based estimate applies its agent share and task mix to worldwide tokens; the alternative estimates each venue separately. Both divide by the workload profiles since October 7, 2026 (coding, research, enterprise workflow, blended on the platform's token mix; the register in B1). Billed hours are small because few products bill in hours yet, and they are a cross-check, not a lower bound: Devin's price covers Windsurf seats as well, and Agentforce's minute a work unit is graded. These methods share workload assumptions, so they are alternative estimates, not independent confirmations and not a confidence interval. The starting point is fixed for the forecast, set on the one-operator profile on October 6, 2026 and kept until the calibration here is reviewed under the rule in A4; the headline is the OpenRouter-based estimate. Measured inputs are marked; the remaining inputs are assumptions.

Tokens a request for OpenRouter’s agent apps over the trailing 30 days, the measured input to the fresh share and the one-operator reference

AGENT APP	TOKENS, 30 DAYS	REQUESTS	TOKENS A REQUEST	FRESH SHARE AT 19K A CALL
Hermes Agent	53.4T	503M	106k	18%
Claude Code	40.2T	360M	112k	17%
Kilo Code	26.4T	280M	94k	20%
Cline	19.6T	160M	123k	16%
pi	13.7T	130M	106k	18%
Codex	11.9T	153M	78k	25%
omp	11.1T	73M	151k	13%
OpenClaw	7.6T	84M	91k	21%
DeepSeek Harness	7.5T	46M	164k	12%
Command Code	4.2T	46M	90k	21%
OpenHands	3.2T	34M	94k	20%
Proto Agent	2.5T	8M	315k	6%
ZCode	2.2T	10M	214k	9%
Cursor	1.1T	8M	139k	14%

Thirty days to 2026-10-22: the platform’s agent apps carried about 108k tokens a request, token-weighted across 14 apps (84% of the top fifty’s tokens), quartiles 92k to 138k; the platform as a whole 82k. From OpenRouter’s app rankings (requests and tokens per app over the trailing 30 days), agent apps as the tracker classifies them; the fresh share is our agents’ 19,000 fresh tokens a call over each app’s context. Refreshed nightly; the tracker draws the series once the record is thirty days long. Measured.

A5 • The human benchmark: 7.0 trillion paid hours a year

We use seven trillion paid human hours a year as the central benchmark: about 3.6 billion people in employment (ILO modelled estimates, November 2025: 3.60 billion in 2025 and 3.64 billion in 2026) at about 2,000 hours each; the ILO’s mean of 41 hours actually worked a week implies closer to 2,100, so seven trillion is a rounded-down benchmark. A defensible range is approximately 6.5 trillion to 7.5 trillion. The central benchmark is held flat in the forecast comparison.

This construction includes self-employment and excludes unpaid household work. Hours actually worked, usual hours and hours paid for are different statistical concepts; the benchmark is an approximation, not a single directly reported ILO total.

Knowledge work is a smaller comparison, about 2.4 trillion hours in 2026 and 2.5 trillion by 2032. Knowledge workers, in our count, are the people whose work is mainly handling information. ILO occupational data put managers, professionals, technicians and clerical workers at 25 to 28% of the employed, about 0.9 to 1.0 billion people in 2026; Gartner counted more than 1 billion knowledge workers in 2019 and Forrester 1.25 billion information workers in 2018 on a wider definition. The model uses 1.128 billion knowledge workers in 2026, rising to 1.179 billion, with an eight-hour day and five-day week: a model input between the two definitions, not a published statistic.

Machine-hour results are annualized rates at the dated reading. They compare duration, not productivity, economic value or jobs displaced.

Appendix B — The machine hours model: the assumption register and Jevons' paradox

B1 · The assumption register

The register lists each input, its source, the judgment applied and the evidence that would change it. The starting adoption rate is an estimate. The 2032 settings are scenario assumptions.

The forecast uses one shared definition of an active hour. Licenses, agent registrations and corporate ambitions do not directly measure it. Survey evidence is kept with its population, question and period.

The token conversion uses three workload profiles. Central processed-token rates are approximately 20 million per hour for coding, 2.5 million for research and 7.5 million for enterprise workflows. The respective ranges are 12 to 30 million, 1 to 8 million and 2 to 14 million. They are judgments informed by traces and benchmarks. The research profile describes long report-writing work; the SimpleQA trace, short single-search questions, runs at about 16 million processed tokens an hour and does not inform that central value.

Some older comparisons combine token use from one benchmark with duration from another. The workflow panel records time, tokens and other observations from the same runs. It includes our coding sessions, enterprise tasks and research questions, plus public run logs from two other coding operators, All Hands AI and METR. Each result stays attached to the workload and evaluation that produced it.

Our workflow trace covers 48 retail and airline tasks. The agent and simulated customer used Claude Haiku 4.5. The agent passed 52%; failed tasks are not assumed to have been completed by a person. The trace reports about 13 million processed tokens, 1.07 million fresh input tokens an hour and 0.32 million written, at about \$4.09 of list-price inference per hour.

The enterprise profile combines that experiment with a separate estimate assembled from production disclosures. Its fresh-input central value is about 1.6 million per hour. The workflow blend is derived from hours, so its token series must use matching definitions.

The useful next step is a broader sample that follows the same task from active work to acceptance, human review, the actual infrastructure bill and repeated paid use. The panel publishes a common contribution format so new observations can be compared.

Our panel includes 6,118 coding runs (300 from All Hands AI’s public evaluation, 5,809 from METR’s public runs and the rest our own sessions) and 48 enterprise tasks. The workflow trace reports 52% accepted, at an estimated \$0.05 of list-price inference per accepted task. The coding panel reports 52 minutes of between-turn waiting per active hour; how much was human review is unmeasured. Our own sessions carry no acceptance record; the two imported logs supply acceptance for the rest, each on its own benchmark, and the pooled share mixes different tasks and operators.

The research agent answered 48 SimpleQA questions with web search: 96% correct under the benchmark’s grading rules, a median of 1 search per question and about 3 active seconds per question. List-price inference cost about \$0.027 per correct answer. These are short factual tasks, so the result does not establish the cost or reliability of a long research report.

The second coding source is All Hands AI’s public OpenHands evaluation on 300 SWE-bench Lite tasks, using Claude 3.5 Sonnet on October 28, 2024. Its published logs show 22 active hours, 42% resolved and about \$1.62 of inference per resolved task. Each column states how many runs support it because the sources record different measures.

A third coding source is METR’s public time-horizon run data: 5,809 runs of five 2025 and 2026 frontier models on its HCAST and RE-Bench software tasks, 82% accepted on the binary score, with each task’s human baseline time kept beside the row (median 11 minutes). The file records each run’s wall-clock span and no per-step timestamps, so these rows add acceptance and human time to the panel but no active hours (see the [METR source data](#)). The import script is not included in this public edition.

Retained paid use remains an open question. The register has 21 dated disclosures and reports covering application creation, deployment, persistence and paid usage. None reports the share of agent-created resources still generating bills after ninety days: the newest creation figures are Supabase’s 70% of new databases and 4 million launches a month (October 2, 2026, from 60% in June) and half of Vercel’s 6 million deployments a day; the platforms’ own claim windows, 72 hours on Neon and 60 minutes on Railway, are graduation gates with no disclosed claim rates; and the only survival-style number is an independent Flathub study, 73% of AI-tagged repositories abandoned within months. We do not yet have invoices linked to agent-created applications over 90 days. That evidence is needed to measure continued paid use.

Operators can contribute comparable observations through the About page, with their names included or withheld.

ASSUMPTION · 2026 → 2029 → 2032	AGENTIC WORKFORCE	MACHINE ECONOMY
Active adoption (share)	12% → 36% → 70%	12% → 62% → 95%

ASSUMPTION · 2026 → 2029 → 2032	AGENTIC WORKFORCE	MACHINE ECONOMY
Human-directed agents/user	1.00 → 4.00 → 16.00	1.00 → 4.00 → 16.00
Human-directed hours/day	0.20 → 0.60 → 1.80	0.20 → 0.60 → 1.80
Autonomous workers/adopter	0.05 → 0.50 → 5.00	0.05 → 1.50 → 45.00
Continuous share	6% → 13% → 25%	6% → 28% → 70%
Continuous hours/day	11.00 → 13.67 → 17.00	11.00 → 15.91 → 23.00
Triggered hours/day	0.55 → 1.39 → 3.50	0.55 → 1.79 → 5.80

Token use per hour for coding, research and enterprise workflows, and the workload mix used to estimate hours

PROFILE	SHARE OF AGENT TOKENS	PROCESSED TOKENS AN HOUR (DENSE TO THIN)	FRESH TOKENS AN HOUR	CADENCE AND DURATION
Coding agents	49%	20M 30M to 12M	1.5M 2.7M to 1.0M	36 to 64 tool calls a task; 3 to 14 minutes a SWE-bench Verified task at the frontier, 35 to 82 minutes a Terminal-Bench task
Research and browsing agents	15%	2.5M 8.0M to 1.0M	1.0M 2.5M to 0.3M	18 to 80 searches a task; 3 to 30 minutes a task; research agents use 4x to 15x a chat's tokens (Anthropic)
Enterprise workflow agents	35%	7.5M 14M to 2.0M	1.6M 7.0M to 0.8M	About 8 to 11 model calls a service conversation (Telepass 8.1; our trace 11), 2 model calls a turn (Agentforce runtime), 1.9 to 4 seconds a call; an Agentforce work unit is one action
Blended, hours = Σ tokens ÷ tokens an hour	100%	7.5M 16M to 2.7M	1.3M 4.4M to 0.6M	the tokens-per-hour assumption used by the OpenRouter estimate and its part of the provider-by-provider estimate; the one-operator profile read 28M processed and 1.4M fresh

Allocated tokens and inferred hours · central assumptions

WORKLOAD	TOKEN SHARE	HOURLY SHARE
CODING AGENTS	49.2%	18.4%
RESEARCH AND BROWSING AGENTS	15.4%	46.3%
ENTERPRISE WORKFLOW AGENTS	35.3%	35.3%

Hour share is each workload’s allocated tokens divided by its assumed tokens per hour, as a share of total inferred hours. A workload with fewer tokens per hour accounts for more hours per token. These are model contributions, not observed shares of worldwide agent work.

Evidence labels describe the source and how it was checked, not independent verification of its claims. The shaded hours range varies tokens per active hour only. It does not cover all uncertainty in worldwide token volume, agent classification or how well OpenRouter represents other platforms.

Coding: tokens per problem from the SWE-rebench leaderboard (111 fresh GitHub problems, May 15 to July 1, 2026, five models) over minutes per task from the vals.ai SWE-bench Verified leaderboard (September 1, 2026), two task sets of the same difficulty class; cross-checked on OpenHands trajectories (1.15M tokens a trajectory over 3,000 runs) and our own Claude Code logs (A2). Research: tokens per task and minutes per task from Google's, OpenAI's and GPT-Researcher's deep-research runs and Agent Laboratory. Enterprise workflow: production telemetry combined, model calls a conversation (Telepass) × tokens a call (AgentCore, Cloudflare, Fin, Salesforce) over agent response time a conversation (Salesforce, AgentCore, Fin), cross-checked on tau-bench episodes; cache share from Datadog and OpenRouter. Fresh shares: 94% to 95% cached in coding traces (SWE-rebench, our logs), about 40% fresh in research, about 55% in workflow, since enterprise traffic caches far less; the enterprise central is the geometric mean of the vendor-derived midpoint and our first-party trace. The long-form research profile remains an assumption until we have comparable research tasks with both token use and active time recorded. Combining tokens from one benchmark and time from another is a cross-study inference, not a matched observation. The mix: OpenRouter's task tags, token-weighted, from the latest daily snapshot (OpenRouter describes them as sampled classifications of platform traffic). The code category is coding; five tags are research; two tags are enterprise workflow (workflow execution, customer support); six chat kinds are dropped; everything else (multi-step planning, tool dispatch, memory extraction, classification, data extraction and transformation, security audit, DevOps, finance) is workload uncertain, since the tag does not say which workflow contains it. The uncertain share is allocated to the three profiles by a stated rule, central in proportion to the classified shares, with three sensitivities: all to coding, all to research and all to enterprise workflow. Latest window ending 2026-10-17. Cross-checks: 46% of first-party API transcripts are computer and mathematical tasks (November 2025); 75% automation (Anthropic Economic Index report: Economic primitives (January 15, 2026)); Programming about 50% of all OpenRouter tokens in late 2025 (a16z and OpenRouter) (OpenRouter and a16z, State of AI: An Empirical 100 Trillion Token Study (arXiv 2601.10088)); Coding 55% of departmental enterprise AI application spend in 2025; customer success 9%, IT operations 10% (Menlo Ventures, 2025: The State of Generative AI in the Enterprise).

Coding agents

2.5M tokens a problem (Fable 5) to 4.3M (Opus 5) on SWE-rebench over 6 to 10 minutes a task on SWE-bench Verified gives 15M to 27M an hour; OpenHands at 1.15M a trajectory gives 5M to 14M; our own logs run about 112M an hour on the execution-interval basis adopted October 7, 2026 (61M on the capped-gap estimate it replaced) at about 420k context a call, and about 26M on OpenRouter's 100k context. Fresh: 5% to 10% of processed in the cached traces (SWE-rebench 95% cached, our logs 96%, the Claude Code cost study about 10%), 1.0M to 2.7M an hour, the top of the range our own one-operator 2.7M.

- Tokens per problem 0.6M to 5.5M across five models (Fable 5 2.52M, Opus 5 4.32M, Grok 4.5 2.43M, GLM-5.2 5.52M, GPT-5.6 Sol 0.61M), 94.9% cached, 111 problems · [Nebius SWE-rebench leaderboard](#), 2026-07-01 · benchmark · Source checked
- Minutes per task 3 to 14 at the frontier (Claude Opus 5 9m37s, Fable 5 5m56s, Grok 4.6 10m04s, GPT-5.6 Sol 3m02s); cost \$0.02 to \$2.05 a task · [vals.ai SWE-bench Verified leaderboard](#) (updated 9/1/2026), 2026-09-01 · benchmark · Source checked
- Terminal-Bench 4.0: 35 minutes to 1h22m a task, \$9.58 to \$30.34 a task · [vals.ai Terminal-Bench 4.0 leaderboard](#), 2026-10-06 · benchmark · Source checked
- 3.44 billion tokens over 3,000 OpenHands trajectories on SWE-bench Verified, about 1.15M a trajectory · [arXiv 2609.24928v1, Trajectory-Aware Benchmark Subset Selection](#) (Ayyad et al.), 2026-09-21 · benchmark · Source checked
- 36 tool calls a trajectory (OpenHands, Claude Sonnet 4, 704 trajectories); 16 to 18 in SWE-Gym and R2E-Gym · [arXiv 2512.12216v2, Training Versatile Coding Agents in Synthetic Environments](#) (Zhu, Gandhi, Neubig), 2026-01-13 · benchmark · Source checked
- 64.3 agent turns a trajectory over 67,074 OpenHands trajectories (Qwen3-Coder) · [Hugging Face dataset card, nebius/SWE-rebench-openhands-trajectories](#), 2025-12-23 · benchmark · Source checked
- Claude Code runs: about 117k cache-read, 12k cache-creation and 715 generated tokens a run (2,848 runs) · [arXiv 2607.12161v1, Token Reduction Is Not Cost Reduction](#), 2026-07 · benchmark · Source checked
- SWE-bench Pro: 2.8M to 3.1M input tokens a task, 7k to 18k output, 616 tasks · [Nilenso, SWE-Bench Pro cost/token/time analysis report](#), 2025-10-13 · vendor disclosure · Source checked
- Ten Claude actions a prompt on average, about four prompts a session, 400,000 sessions · [Anthropic, How Claude Code is used in practice](#), 2026-04 · vendor disclosure · Source checked

Research and browsing agents

Gemini Deep Research about 310k tokens in 5 to 15 minutes (1.9M an hour); o3 deep research 1.26M tokens a call over 5 to 30 minutes (2.5M to 15M); GPT-Researcher 375k tokens in 57 minutes (0.4M); Agent Laboratory \$2.33 and 19 minutes a paper (about 2M). Fresh about 40%: fetched pages are new input, the accumulating context is cached.

- Gemini Deep Research: about 250k input and 60k output tokens, about 80 searches, 5 to 15 minutes a standard task (Google documentation as relayed) · [tokencost.app: Gemini Deep Research agent cost](#) (secondary, cites Google documentation), 2026-04-24 · Secondary report · underlying documentation cited · Reported via secondary source
- o3 deep research: 1,260,699 tokens for one API call (developer report) · [OpenAI Developer Community thread](#) (user report), 2025-06-28 · Developer report · anecdotal · Not independently verified

- GPT-Researcher: 375,400 tokens and 57 minutes a report, 29 nodes explored · [arXiv 2608.08389](#), Not Worth Another Token: Marginal Value Estimation for Efficient Deep Research Agents, 2026-08-09 · benchmark · Source checked
- Agent Laboratory: \$2.33 and 19.4 minutes a paper on gpt-4o · [arXiv 2501.04227](#), Agent Laboratory: Using LLM Agents as Research Assistants, 2025-01-08 · benchmark · Source checked
- Deep Research Bench: \$0.46 to \$0.53 and 140 to 262 seconds a task (Claude Opus 4.5, 4.6, Sonnet 4.6) · [FutureSearch evals leaderboards \(Deep Research Bench\)](#), 2026-06-10 · benchmark · Source checked
- BrowseComp-Plus: 6 to 24 search calls a query by model; GPT-5 \$0.43 a query · [arXiv 2508.06600](#), BrowseComp-Plus, 2025-08-08 · benchmark · Source checked
- Research agents use about 4x a chat's tokens, multi-agent research 15x; 3 to 15 tool calls an agent · [Anthropic Engineering: How we built our multi-agent research system](#), 2025-06-13 · vendor disclosure · Source checked
- Claude Advanced Research: 5 to 45 minutes a task, most reports in 5 to 15 · [Anthropic/Claude blog: Claude can now connect to your world \(Integrations and Advanced Research\)](#), 2025-05-01 · vendor disclosure · Source checked
- Perplexity deep research: 18 searches and 20,000 citation tokens a request, medians · [Perplexity API pricing docs](#), 2026-10-07 · vendor disclosure · Source checked

Enterprise workflow agents

Two sources, graded together. Production telemetry combined, since no vendor publishes the profile whole: about 8.1 model calls a service conversation (Telepass) at about 4,500 tokens a call (AgentCore, Cloudflare, Fin, Salesforce) over 20 to 40 seconds of agent response time (Salesforce, AgentCore, Fin) gives 3 to 6.5 million tokens an hour, about 4.5 million at the midpoint, with about 55% fresh because enterprise traffic caches little (Datadog). Our own first-party trace, October 7, 2026 ([scripts/measure-workflow-agent.py](#), [workflow-trace.json](#)): a tool-calling service agent on 48 tau-bench retail and airline tasks, Claude Haiku 4.5 as agent and simulated customer, policy, tools and conversation cached: 11 model calls over 21 seconds of agent time a task at the median, 1.9 seconds a call, 13.0 million tokens processed an hour (quartiles 10.7 to 14.4), 1.07 million fresh input (new input and cache writes, quartiles 0.75 to 1.23) and 0.32 million written, reported separately since October 7, 2026 (fresh had included written tokens), cache share 89%; the agent's own CPU was 5% of its time and tool execution 0.2%. The trace is denser than production on processed tokens because a fast model with full caching answers in 2 seconds where production agents take 2.5 to 7, and thinner on fresh tokens because production caches far less. The central is the geometric mean of the two, about 7.5 million processed and 1.6 million fresh input an hour (production about 2.5 million at 55% fresh of 4.5 million, the trace 1.07 million); the band runs from the vendor-derived thin end to the trace's dense end. Active execution is the agent's response time, model inference included, the same basis as our own coding logs; a CPU-time basis would read twenty times denser. The trace re-runs monthly in CI and the profile is re-graded if two months sit outside the band.

- First-party trace: 48 tau-bench tasks, 11 model calls and 21 seconds of agent time a task (median), 13.0M tokens processed, 1.07M fresh input and 0.32M written an hour, 89% cached, 1.9 s a call, pass rate 52% · [Skycatcher workflow trace: Claude Haiku 4.5 agent and](#)

customer. Method summarized in Appendix B1; the script and raw trace are not included in this public edition., 2026-10-07 · first-party trace · Measured by [Skycatcher](#)

- Telepass on Agentforce: 1.4 million LLM calls a month over about 40,000 conversations a week, about 8.1 calls a conversation; 87% resolved autonomously · [Salesforce customer story, Telepass, 2026-01-15](#) · case study · Source checked
- Agentforce runtime: 2 serial LLM calls a turn before streaming (from 4), a 70% latency reduction · [Salesforce Engineering, How we rearchitected the Agentforce runtime, 2026-01-26](#) · production telemetry · Source checked
- Salesforce: 28.6 trillion tokens over 3.8 billion work units to date, at most about 7,500 tokens a work unit · [Salesforce Q1 FY27 earnings release \(SEC 8-K\), 2026-05-27](#) · production telemetry · Source checked
- Cloudflare's internal code-review agent: 24.77 billion tokens over 5.47 million requests in 30 days, about 4,530 tokens a request · [Cloudflare blog, The AI engineering stack we built internally, 2026-04-20](#) · production telemetry · Source checked
- AgentCore Observability sample: 15,700 tokens over 6 traces, 3.76 seconds a trace · [AWS Machine Learning Blog, Optimizing production agents with AgentCore Observability, 2026-07-31](#) · vendor worked example · Source checked
- Intercom Fin: about 6,000 input and 260 output tokens a call; 1,700 output tokens with reasoning · [Fin AI Research, Think Fast: Reasoning at 3ms a Token, 2025-07-18](#) · production telemetry · Source checked
- Intercom Fin: 7-second median time to first token in production · [Fin AI Research, Slower Feels Smarter?, 2025-03-14](#) · production telemetry · Source checked
- Sierra: 10 to 15 model invocations a turn, a third frontier models, under 2 seconds · [ZenML LLMOps Database on Sierra's constellation of models, 2025-12-03](#) · Secondary report · Reported via secondary source
- Salesforce: under 5 seconds a turn feels fine, 6 to 10 acceptable · [Salesforce Help, Troubleshooting Agentforce latency, 2026-08-20](#) · vendor worked example · Source checked
- Datadog: 28% of LLM spans show any cached-read tokens; 69% of input tokens are system prompt; 59% of agentic requests make one service call · [Datadog, State of AI Engineering 2026, 2026-07-30](#) · production telemetry · Source checked
- OpenRouter: nearly 70% of agent token usage from cached prompts (agentic traffic, coding-heavy) · [The Decoder citing OpenRouter \(Peter Walker\), 2026-08-23](#) · Secondary report · Reported via secondary source
- Vercel AI Gateway: under 8% of runtime hours were CPU work, the rest waiting on model providers · [Vercel blog, How AI Gateway runs on Fluid compute, 2025-11-06](#) · production telemetry · Source checked
- Production coding-agent traces: 20% of session lifetime in LLM and tool execution at the median, 35,037 sessions · [Chang et al., From LLM Inference to Agentic Workloads \(arXiv 2608.15127\), 2026-08-15](#) · production telemetry · Source checked
- tau-bench: \$0.38 an episode agent-side on gpt-4o, input 95.9% of the price, at most 30 actions a task · [arXiv 2406.12045, tau-bench, 2024-06-30](#) · benchmark · Source checked

- Agentforce: 3.2 billion Agentic Work Units in the quarter to July 31, 2026, one discrete action each · [Salesforce Q2 FY27 earnings press release, 2026-07-31](#) · production telemetry · Source checked
- Klarna: errands resolved in under 2 minutes against 11; 2.3 million conversations in the first month · [Klarna press release as republished by VKTR, 2024-02-27](#) · Secondary report · Reported via secondary source

The central token rates are judgments within the sourced ranges. More tokens per hour imply fewer hours for a given volume; fewer tokens imply more hours. The enterprise profile combines production disclosures with the separate service-agent benchmark in B1. The matched panel includes our coding, service and research runs, plus public run logs from All Hands AI (OpenHands) and METR. Vendor disclosures provide context for retained paid use; a matched ninety-day billing record is still missing. Sources are dated October 7, 2026.

Time, tokens and outcomes from coding, enterprise and research runs, including public run logs from two other coding operators, All Hands AI and METR; each column states how many runs it is graded on, and missing observations are shown as not graded.

Coding panel: 6,109 acceptance-graded runs, 309 timed runs, and nine first-party fresh-token/hour rows. METR has no active-hour measure; OpenHands uses capped gaps; first-party sessions classify execution intervals. These timing cohorts are not interchangeable.

SOURCE	WORKLOAD	RUNS	DATES	ACTIVE HOURS	ACCEPTED	LIST COST AN HOUR · PER ACCEPTED TASK	BETWEEN-TURN WAITING MIN/HOUR
All Hands AI (OpenHands), public evaluation log SWE-bench_Lite-test · CodeActAgent · anthropic/claude-3-5-sonnet-20241022	Coding agents	300	October 28, 2024	21.9	42% of 300	\$9.22 · \$1.62	not graded
METR, public time-horizon run data Time Horizon 1.1 runs · HCAST and RE-Bench software tasks · Gemini 3 Pro, GPT-5.1-Codex-Max (Inspect), Claude Opus 4.6 (Inspect), Claude Opus 4.5 (Inspect), GPT-5.2	Coding agents	5,809	November 26, 2025 to February 14, 2026	wall-clock only, 61,639 h of spans	82% of 5809	not graded · not graded	not graded · human baseline 11 min a task
Skycatcher, first party Claude Code session	Coding agents	9	July 15, 2026 to October 8, 2026	97.9	not graded	not graded · not graded	52
Skycatcher, first party SimpleQA	Research agents	48	October 7, 2026	0.04	96% of 48	\$30.85 · \$0.027	0
Skycatcher, first party tau-bench	Enterprise workflow agents	48	October 7, 2026	0.33	52% of 48	\$4.09 · \$0.054	0

WORKLOAD	RUNS · OPERATORS	ACTIVE HOURS	TOOLS RUNNING	PEAK CONCURRENCY	TOOL ERRORS AN HOUR	BETWEEN-TURN WAITING MIN/HOUR	ACCEPTED	LIST COST PER ACCEPTED TASK	FRESH TOKENS AN HOUR
Coding agents 3 operators: first party and 2 public evaluation logs	6118 · 3	See separate cohorts above	34%	1	6.9	52	80% of 6109	not graded	6.4M (9 runs)
Enterprise workflow agents first party, one benchmark	48 · 1	0.33	0%	1	not graded	0	52% of 48	\$0.05	1.1M
Research agents first party, one benchmark	48 · 1	0.04	not graded	1	not graded	0	96% of 48	\$0.03	16.3M

Coding agents: Different timing cohorts; the pooled hours are descriptive, not a common calibrated meter. 5,809 METR runs contribute acceptance but no active time. 300 OpenHands runs use capped inter-event gaps; nine first-party sessions classify execution intervals. Only those nine calibrate fresh tokens per active hour. Between-turn waiting is not observed human review. Legacy first-party list costs are withdrawn where the original token/cache split cannot be recovered. Sources: our own sessions carry a person’s review time and tool errors but no acceptance record, the imported benchmark log carries acceptance and cost but no tokens or tool errors; each column states how many runs it is graded on. Between-turn waiting is bounded at ten minutes a gap; it does not observe human review. No acceptance record exists in the session logs, so the accepted share is not graded here. Cost is list price where every call in the session is on our price table; sessions on models outside it carry no cost.

Enterprise workflow agents: The tau-bench retail and airline tasks with a simulated customer: human review is zero by construction and the failed share is tasks the agent did not complete; subsequent human handling was not measured. Retries are not instrumented in the trace.

Research agents: A web-searching agent on a fixed sample of SimpleQA, short-answer factual questions graded against a reference by the benchmark’s own rubric: a research task of one or two searches, not a long report. The searches run inside the call, so tool time is not separable from the model’s; the not-attempted and incorrect shares are tasks the agent did not complete; subsequent human handling was not measured. The sample is every ninetieth question, so it is reproducible.

Imported public logs: METR, public time-horizon run data ([the log](#), public repository, no license stated, February 14, 2026); All Hands AI (OpenHands), public evaluation log ([the log](#), MIT, October 28, 2024). The OpenHands timing proxy sums inter-event gaps capped at ten minutes; unlike first-party execution-interval classification it does not distinguish model work from all intervening time, calls and cost from the run’s cost record, the accepted outcome from the benchmark report; tokens and tool errors are not in that log. The METR file carries each run’s wall-clock span and no per-step timestamps, so its rows add acceptance, tokens and the task’s human baseline time but no active hours.

Retained paid usage, read from vendor disclosures by stage

STAGE	PLATFORM · DATE	AS DISCLOSED	STATUS
Created by agents	Neon (Databricks) January 27, 2026	80% of databases on Neon are created by agents.	Source checked

STAGE	PLATFORM · DATE	AS DISCLOSED	STATUS
Created by agents	Supabase June 4, 2026	In the past year, database launches on Supabase have grown 600%. More than 60% of new databases are launched by some sort of AI tool.	Source checked
Created by agents	Databricks Lakebase June 16, 2026	Lakebase handles 12 million database launches per day across the platform; the release does not say what share agents create.	Source checked
Created by agents	Vercel April 13, 2026	30% of the applications on the platform originated from agents (Guillermo Rauch to TechCrunch).	Reported via secondary source
Built to deployed	Vercel June 17, 2026	Coding agents trigger more than half of Vercel's deployments, from under 3% six months earlier.	Reported via secondary source
Created by agents	Supabase July 23, 2026	We can measure 60% of databases get launched by an agent, but more likely it's like 90% and it's in the millions every month (Paul Copplestone, as reported).	Reported via secondary source
Created by agents	Microsoft (GitHub) July 29, 2026	One in three pull requests on GitHub now involves an agent (fiscal 2026 fourth-quarter call).	Source checked
Created by agents	Supabase October 2, 2026	Supabase is now adding more than 1M users and 4M databases per month with 70% of new databases created by agents or AI-driven tools.	Source checked
Created by agents	Supabase October 2, 2026	Supabase is already launching over one million databases per week; agents are spinning up millions of databases to power the prototypes, explorations, dashboards and apps.	Source checked
Built to deployed	Emergent on MongoDB Atlas November 30, 2025	In the four months since Emergent Labs' public launch, the platform has supported the creation of nearly 2 million applications, with 50,000 successfully deployed and serving users across 180 countries.	Source checked
Built to deployed	Emergent on MongoDB Atlas November 30, 2025	The deployment rate has doubled from 10% to 20% in three months.	Source checked
Built to deployed	Emergent on MongoDB Atlas November 30, 2025	About 2.5% of the applications built in the platform's first four months were deployed: 50,000 of nearly 2 million.	Calculated
Built to deployed	Vercel July 6, 2026	The company currently sees 6 million deployments a day, half of them triggered by coding agents (Guillermo Rauch to TechCrunch).	Reported via secondary source
Kept running	Flathub (independent study) July 18, 2026	Of 120 repositories tagged as AI-generated, 88 (73%) were abandoned, with no activity in two to three months; a third-party study, not a vendor figure.	Reported via secondary source
Kept running	Convex August 4, 2026	Millions of Convex instances are now running in production.	Source checked

STAGE	PLATFORM · DATE	AS DISCLOSED	STATUS
Kept running	Lovable August 12, 2026	People have created more than 60 million projects; Lovable-built apps see over 900 million visits every month; more than one-third of builders who plan to monetize are already earning revenue.	<u>Source checked</u>
Kept running	Railway August 20, 2026	Anonymous provisioning: you have 60 minutes to build and then claim; Railway publishes no claim rate.	<u>Source checked</u>
Kept running	Neon October 7, 2026	A popular developer platform managed over 300k Postgres instances on Neon with only a single engineer. Tens of thousands of projects can spin up and idle as users create apps; scale-to-zero ensures that idle databases cost you nothing while remaining instantly accessible.	<u>Source checked</u>
Kept running	Neon October 7, 2026	A claimable Postgres database created for an agent's user expires after 72 hours unless it is claimed; Neon publishes no claim rate.	<u>Source checked</u>
Paid usage	Supabase June 4, 2026	Supabase for Platforms, the product that hosts other companies' app builders, has been the fastest growing product over the last year.	<u>Source checked</u>
Paid usage	Databricks Lakebase August 13, 2026	Exceeding \$100M revenue run-rate for Lakebase; the release does not attribute it to agents.	<u>Source checked</u>

Three stages, disclosed unevenly. Creation is disclosed by several platforms and keeps rising (Supabase 60% in June 2026, 70% in October). Graduation (built to deployed) is disclosed by one, Emergent on MongoDB Atlas. Paid usage is disclosed at the vendor level, never per artifact. We had not found a platform disclosure of how many agent-created resources were still billing 30, 60 or 90 days after creation as of October 7, 2026, which is the figure the panel needs (searched October 7, 2026 across Supabase, Neon, Databricks, Convex, MongoDB, Vercel, Cloudflare, Replit, Lovable, Bolt, Railway, Render, Fly.io, PlanetScale, Turso, Upstash, Firebase, AWS and Microsoft). The platforms' own claim windows (Neon 72 hours, Railway 60 minutes) are graduation gates with no disclosed claim rates, and the only survival-style number is an independent study of AI-tagged repositories on Flathub.

Retained paid usage: read from vendor disclosures; no operator ledger yet. Paid backend consumption that is still billing 90 days after the run that created it, read from an operator's billing matched to run ids. A billing export in docs/research/billing/*.csv with the columns date, provider, service, amount_usd, created_by, run_date. The first-party ledger needs the SpacetimeDB Maincloud, Vercel and GitHub Actions statements for the services the agents' work runs on, matched to the sessions that built them; the Anthropic cost report needs an admin key, which the research key is not. Built October 23, 2026 by build-workflow-panel.py from the same logs as A2 and the trace in this appendix; one row per run, no content read. Contributions carry the same fields (workload, date, activeHours, toolWaitShare, humanReviewMinutes, concurrencyPeak, retries, calls, tokensProcessed, tokensFresh, tokensOut, accepted, listCostUsd): One row per run, the fields above from the same run, no content. Write to us from the About page; rows are published with the operator named or anonymized as the operator chooses.

B2 · Jevons' paradox: the illustration, the precedents and the elasticity condition

Total spending is price multiplied by quantity. A tenfold price reduction requires more than a tenfold increase in comparable usage for spending to rise.

At an assumed demand elasticity of 1.3, a tenfold reduction produces about twenty times the use. At 0.7, it produces about five times. These are illustrations of the condition, not measured forecasts.

Computing, storage and bandwidth offer precedents for large increases in use as costs fall. They do not supply a coefficient for agent demand. Better AI can also reduce the time and resources needed for a fixed task. Our thesis requires enough additional useful work to outweigh those savings.

01 cost per task falls from \$10 to \$1, tasks performed rise from 100 to 2,000, total spending rises from \$1,000 to \$2,000

[Skycatcher](#) illustration based on the source below. Conceptual summary, not a measured quantitative series.

[Skycatcher](#) illustration of Jevons' paradox. Hypothetical numbers, not a forecast: at 90% lower cost, usage must rise above 10x for total spending to increase.

1 Lower cost

Makes some previously uneconomic tasks worth attempting.

2 Demand response

Determines whether more tasks outweigh fewer resources per task.

3 Paid capture

Determines which suppliers retain the spending.

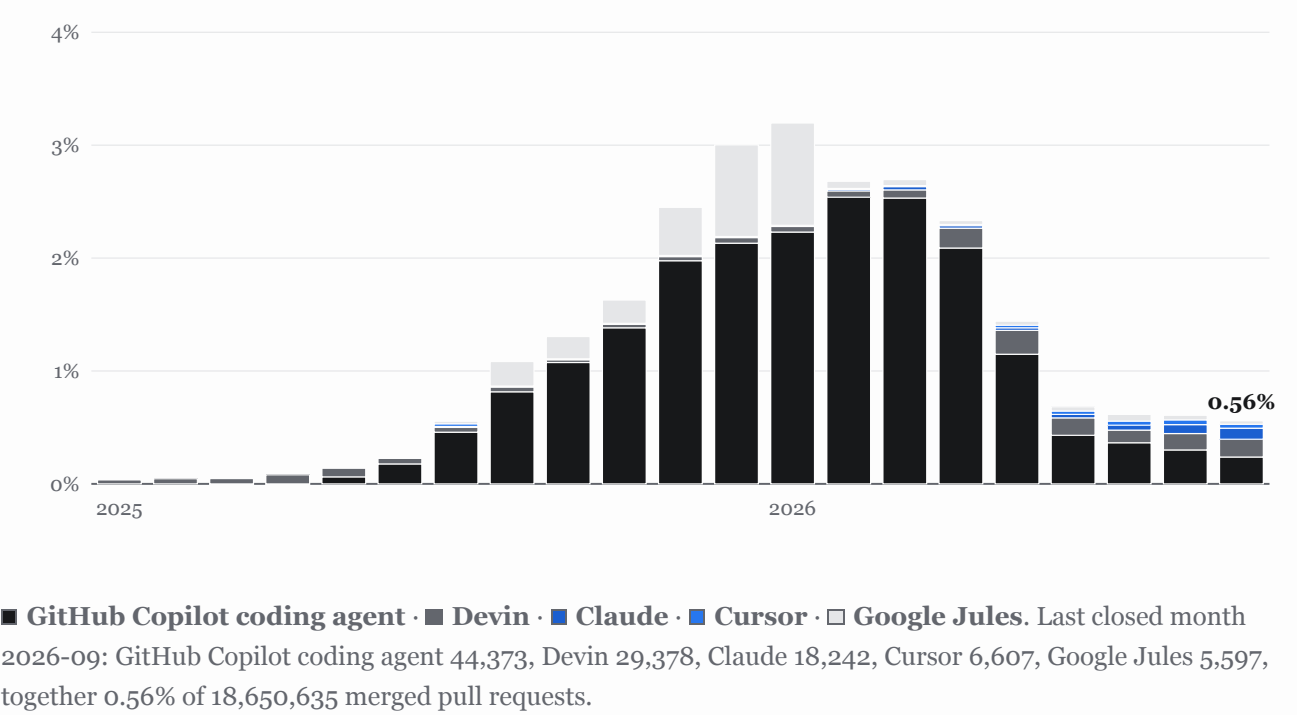
B3 · Displacement: machine work replacing human work where both can be counted

Public GitHub records identify some code changes submitted by named agents. They miss agent work submitted under a human identity. Merged changes indicate acceptance into a codebase, not a universal quality standard.

Stack Overflow records show how many questions people ask publicly. The decline is consistent with a shift toward model assistance, but it does not establish the amount of human work displaced.

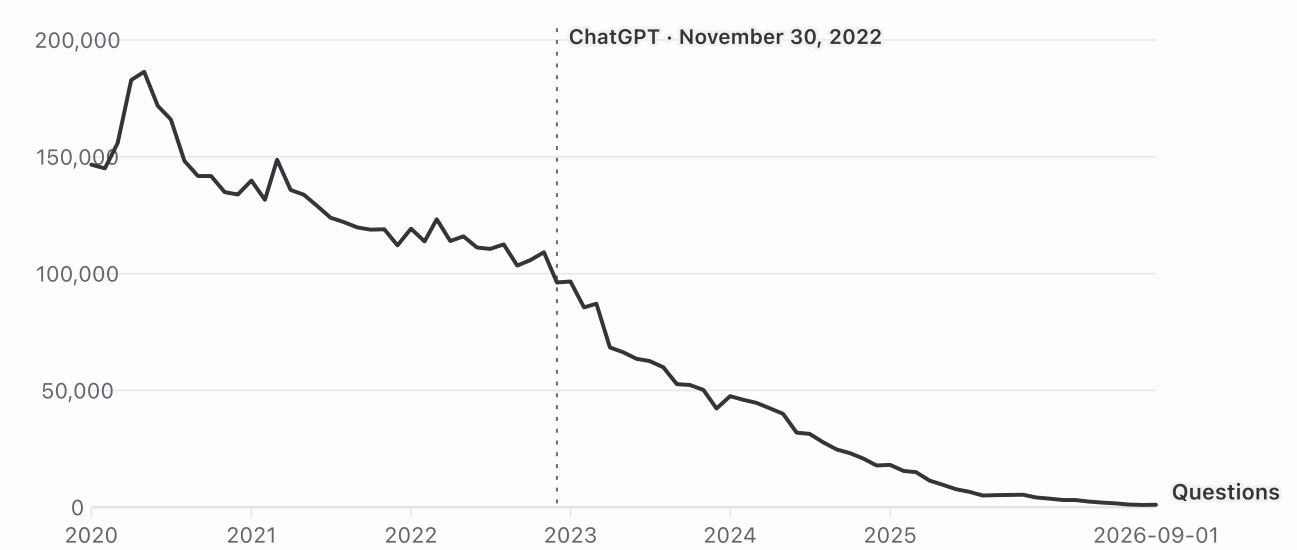
We show these series together as evidence of changing work patterns. They cover different activities and cannot be subtracted to calculate jobs or hours replaced.

Pull requests merged by coding agents, as a share of all merged public pull requests, by agent



Questions asked on Stack Overflow a month, since January 2020, from the Stack Exchange API (content CC BY-SA 4.0)

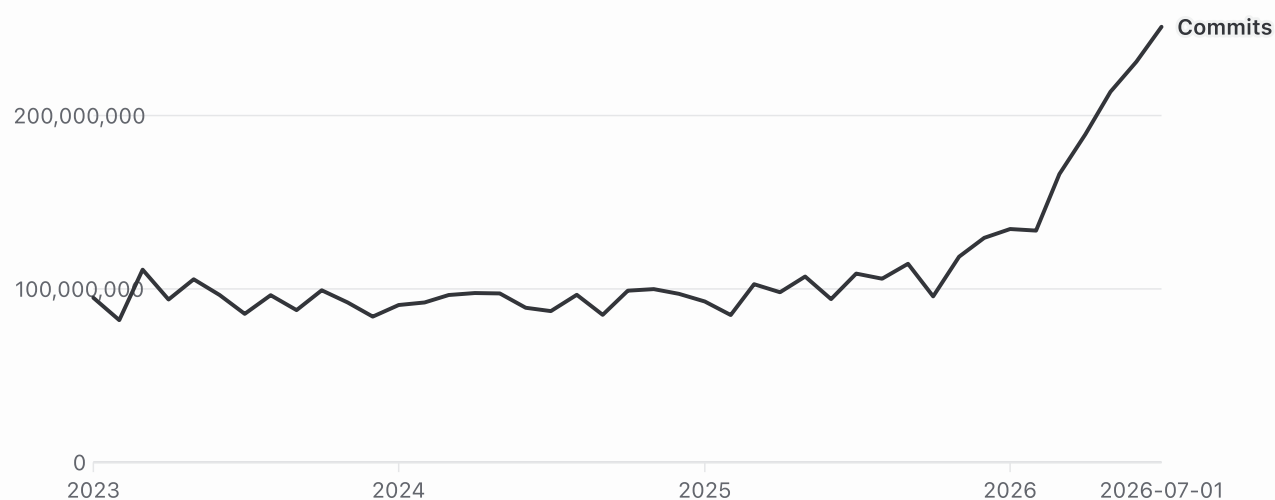
QUESTIONS ASKED A MONTH



1,134 questions in 2026-09, against 146,586 in 2020-01 (-99%). Stack Exchange API, closed months only, stored at first read and never revised.

Public commits a month on GitHub, the weaker production series: a search index that drifts

PUBLIC COMMITS A MONTH



251,217,359 public commits in 2026-07, against 94,930,831 in 2023-01. Public repositories only; the running month is never drawn. Commit counts come from a search index that drifts, so a month that moves more than 50% against the one before is left blank until a later run reproduces it. The two strong series, merged pull requests and new repositories, are on the tracker's production section. Measured.

B4 · The proxies: the dated disclosures behind the inputs

The source tables collect dated observations from Census, Gallup, Ramp, Anthropic, OpenAI, Cloudflare, GitHub, Cognition, Salesforce and other providers.

Comparable observations form a time series. Changes in definition break the line. One-off statements remain table entries rather than becoming a smooth trend through interpolation.

Each entry says what was counted, who was observed and which model assumption it informs. Frontier users show what is possible in an intensive setting. They do not establish the average user's behavior.

The disclosures behind the proxies that have no comparable series: one table per source, each row dated and sourced

OpenAI frontier anchors

What OpenAI's own disclosures say about agent hours in its research organization and among intensive internal users.

DATE	AS STATED	STATUS	SOURCE
August 15, 2026	In terms of a standard 8 hour workday, as of mid-August, in total, the research organization uses 3.1 agent-workdays of effort for every workday of human labor.	Source checked	openai.com
June 30, 2026	By June 2026, OpenAI's internal daily active Codex users at the 99th percentile regularly generated more than 60 hours of agent turns per day across parallel agents.	Source checked	openai.com

Informs: Human-directed agents and hours per user; context for the 2032 assumptions. Agent workdays per human workday across OpenAI's research organization (September 6, 2026 post) and agent hours among its internal daily active Codex users at the 99th percentile (June 25, 2026 post). Some intensive users already exceed the human-directed component of our 2032 assumptions. That does not validate our autonomous-agent count or working-time assumptions.

Cloudflare Radar: agents and crawlers on the web

Cloudflare's disclosures on AI agent and crawler traffic across the fifth of the web it fronts.

DATE	AS STATED	STATUS	SOURCE
July 31, 2024	[training share of AI bot activity] 72% a year ago [July 2024]	Source checked	blog.cloudflare.com
April 30, 2025	April 2025 alone was up 32% compared with May 2024 [AI crawling traffic]	Source checked	blog.cloudflare.com
May 31, 2025	In Spring 2025, 22% of the crawler requests we saw were for AI training	Source checked	blog.cloudflare.com
June 30, 2025	In June 2025, crawling traffic was still 24% higher year-over-year	Source checked	blog.cloudflare.com
July 31, 2025	by July [2025] the increase [in AI crawling traffic, YoY] was down to just 4%	Source checked	blog.cloudflare.com
July 31, 2025	Training now drives nearly 80% of AI bot activity, up from 72% a year ago	Source checked	blog.cloudflare.com
December 2, 2025	early December 2025 HTML requests: humans 47%, non-AI bots 44%, Googlebot 4.5%, AI bots 4.2% (automated share = 100 - 47)	Source checked	blog.cloudflare.com
December 16, 2025	AI user-action crawling 'increasing by over 15x in 2025' (Jan-early Dec 2025); growth 'over 21x from January through early December' (converted: 15x = +1,400%)	Source checked	blog.cloudflare.com
May 31, 2026	Cloudflare's May 2026 data shows that 51.8% of AI crawler requests are for training purposes, and only 9.3% are for search.	Reported via secondary source	workos.com
June 3, 2026	Cloudflare Radar ... automated requests at 57.5% of HTML web traffic versus 42.5% from humans (announced by Matthew Prince on X, June 3, 2026)	Reported via secondary source	techtimes.com
June 5, 2026	seven-day window ending June 5, 2026: bots 57.4% of web traffic to HTML content, humans 42.6%	Reported via secondary source	ppc.land
June 30, 2026	52% of crawler requests are now for AI training as of June 2026, up from 22% in Spring 2025	Source checked	blog.cloudflare.com

DATE	AS STATED	STATUS	SOURCE
September 30, 2026	Over the past year, daily requests from AI agents on our network grew by more than 1,700%.	Source checked	blog.cloudflare.com

Informs: Autonomous and browsing agents’ actions on the web; growth of agent activity. Growth rates and shares with different denominators, so a table rather than a line: agent requests, crawler requests, the automated share of HTML requests and the training share of crawler requests. The weekly Radar series behind them, pulled from the API since October 6, 2026, is the card above.

GitHub: agent-opened pull requests

Pull requests and reviews on GitHub that an agent opened or joined.

DATE	AS STATED	STATUS	SOURCE
September 30, 2025	Pull requests opened by AI coding agents on GitHub jumped from roughly 4 million in September 2025 (The Information)	Reported via secondary source	danilchenko.dev
March 31, 2026	to more than 17 million in March 2026, a 4x increase in just six months (The Information)	Reported via secondary source	ai2.work
April 7, 2026	GitHub Copilot's newly GA coding agent generates roughly 1.2 million pull requests per month	Reported via secondary source	agentmarketcap.ai
May 7, 2026	More than one in five code reviews on GitHub now involve an agent.	Source checked	github.blog

Informs: Production output; growth of agent work in code. The 4 and 17 million figures are The Information’s, carried by secondary accounts; the 1.2 million is Copilot’s coding agent alone; GitHub itself states more than one in five code reviews involve an agent (May 7, 2026). The production section’s Figure 1.6 carries GitHub’s own monthly counts.

Model Context Protocol: servers

Counts of MCP servers, as active public servers (Anthropic) or official-registry entries.

DATE	AS STATED	STATUS	SOURCE
November 25, 2025	the central index for all available MCP servers that now has close to two thousand entries since its announcement in September. That's a 407% growth	Source checked	blog.modelcontextprotocol.io
December 9, 2025	There are now more than 10,000 active public MCP servers	Source checked	anthropic.com
May 31, 2026	about 9,650 servers in May 2026 (official registry)	Reported via secondary source	dev.to
August 22, 2026	24,240 MCP Servers In The Official Registry (as of August 22, 2026)	Reported via secondary source	jwatte.com

DATE	AS STATED	STATUS	SOURCE
September 10, 2026	30,375 unique servers (99,114 server-and-version records) pulled from the official registry API as of September 10, 2026	Reported via secondary source	dev.to

Informs: Tool-call infrastructure; growth only. Two measures: Anthropic’s active public servers (10,000 in December 2025) and the official registry’s entries (about 2,000 in November 2025 to 30,375 in September 2026).

Salesforce Agentforce: deals since launch

Cumulative Agentforce deals, paid where disclosed.

DATE	AS STATED	STATUS	SOURCE
April 30, 2025	closed over 8,000 deals since launching Agentforce, of which half are paid (Q1 FY26)	Source checked	salesforce.com
July 31, 2025	closed over 12,500 deals since launching Agentforce, of which over 6,000 are paid (Q2 FY26)	Source checked	salesforce.com
October 31, 2025	over 18,500 Agentforce deals since launch ... over 9,500 paid deals, up 50% quarter-over-quarter (Q3 FY26)	Source checked	salesforce.com
January 31, 2026	Closed over 29,000 Agentforce deals since launch, up 50% Q/Q (Q4 FY26; paid/unpaid split not stated)	Source checked	sec.gov

Informs: Autonomous agents sold; adoption at the firm. Paid deals to Q3 FY26 (9,500); Q4 FY26 gives only total deals (29,000); from Q1 FY27 Salesforce reports work units instead of deals.

Agentic commerce: x402 and the Machine Payments Protocol

Payments agents settle on their own over HTTP-402 protocols, as Coinbase discloses or explorers index.

DATE	AS STATED	STATUS	SOURCE
October 26, 2025	x402: cumulative transactions passed 1.446 million (x402scan)	Reported via secondary source	panews.io
March 31, 2026	x402: 100M+ total payments processed as of Q1 2026 (Coinbase earnings deck, May 7, 2026)	Source checked	sec.gov
March 31, 2026	MPP: about 31,100 transactions and 671 registered agents two weeks after the March 18, 2026 launch (MPPScan)	Reported via secondary source	formo.so
April 21, 2026	x402: 69,000 agents processed more than 165 million transactions, \$50 million in volume, \$0.31 average	Reported via secondary source	finance.yahoo.com
July 31, 2026	MPP: about 115,000 transactions and \$25,000 settled in its first weeks (Artemis Analytics, via Juniper Research)	Reported via secondary source	juniperresearch.com
August 27, 2026	x402: over 205 million transactions, \$53 million in volume, 200,000 sellers (Coinbase)	Reported via secondary source	odaily.news
October 6, 2026	x402: 75.41 million transactions and \$24.24 million in the trailing 30 days (x402.org live counter)	Source checked	x402.org

DATE	AS STATED	STATUS	SOURCE
October 6, 2026	MPP: 4 million transactions all time, \$434,000 in volume, 224,000 agents (MPPScan)	Source checked	mppscan.com

Informs: Autonomous agents acting economically; an early signal for the Machine Economy, too small for a level. Counts mix cumulative totals and trailing 30-day windows, and dashboards aggregate chains differently, so a table rather than a line. Third parties have estimated that a large share of x402 activity is test or artificial; the rows report counts as stated. Stripe has disclosed no MPP counts of its own.

HUMAN Security: agentic traffic across its customer base

HUMAN Security’s measurement of traffic from AI agents and agentic browsers across the sites it protects, as growth rates and shares.

DATE	AS STATED	STATUS	SOURCE
December 31, 2025	Traffic from AI agents and agentic browsers grew 7,851% year over year in 2025, from a very low base	Source checked	humansecurity.com
December 31, 2025	Agents and agentic browsers were 1.7% of AI-driven traffic in December 2025; training crawlers 74%, scrapers 24%	Source checked	humansecurity.com
December 31, 2025	Automated traffic grew 23.5% in 2025 against 3.1% for human traffic	Source checked	humansecurity.com
August 31, 2026	Agentic traffic grew 27% month over month to a new high in August 2026; Codex debuted at 7.3% of it	Reported via secondary source	humansecurity.com

Informs: Growth of agent actions on the web; an annual cross-check on Cloudflare. The annual report is verified from the PDF; the monthly series since April 2026 is read from secondary accounts because the blog blocks fetches. Shares in the monthly series are shares of agentic traffic, not of all traffic.

Knowledge workers: the population the model counts

Outside counts of knowledge workers worldwide, against the model’s 1,128 million in 2026 rising to 1,179 million in 2032.

DATE	AS STATED	STATUS	SOURCE
December 31, 2019	Gartner: the world passed one billion knowledge workers in 2019, modeled from per-country worker counts	Source checked	blogs.gartner.com
December 31, 2023	Gartner forecast (September 2019): 1.14 billion knowledge workers by 2023, four fifths of the growth in the emerging world	Source checked	gartner.com
July 2, 2025	IDC: 888 million knowledge workers and developers worldwide in 2025, forecast to reach 1.2 billion in 2029	Source checked	my.idc.com

Informs: The people term; the model’s count sits inside the published band of 0.9 to 1.2 billion. IDC’s figure is from the public abstract of a paid census and includes developers; Gartner’s are a 2019 estimate and a 2019 forecast for 2023, with no newer count found.

Agent-built infrastructure: created, deployed, kept, paid

What the platforms that host agent-built software disclose about the stock those agents create: the share of new resources agents create, the share of what is built that graduates to a deployment, the fleets that persist, and the paid usage that follows. Each row keeps its own denominator, as stated.

DATE	AS STATED	STATUS	SOURCE
January 27, 2026	80% of databases on Neon are created by agents.	Source checked	databricks.com
June 4, 2026	In the past year, database launches on Supabase have grown 600%. More than 60% of new databases are launched by some sort of AI tool.	Source checked	supabase.com
June 16, 2026	Lakebase handles 12 million database launches per day across the platform; the release does not say what share agents create.	Source checked	databricks.com
April 13, 2026	30% of the applications on the platform originated from agents (Guillermo Rauch to TechCrunch).	Reported via secondary source	techcrunch.com
June 17, 2026	Coding agents trigger more than half of Vercel's deployments, from under 3% six months earlier.	Reported via secondary source	finance.yahoo.com
July 23, 2026	We can measure 60% of databases get launched by an agent, but more likely it's like 90% and it's in the millions every month (Paul Copplestone, as reported).	Reported via secondary source	finance.biggo.com
July 29, 2026	One in three pull requests on GitHub now involves an agent (fiscal 2026 fourth-quarter call).	Source checked	microsoft.com
October 2, 2026	Supabase is now adding more than 1M users and 4M databases per month with 70% of new databases created by agents or AI-driven tools.	Source checked	prnewswire.com
October 2, 2026	Supabase is already launching over one million databases per week; agents are spinning up millions of databases to power the prototypes, explorations, dashboards and apps.	Source checked	supabase.com
November 30, 2025	In the four months since Emergent Labs' public launch, the platform has supported the creation of nearly 2 million applications, with 50,000 successfully deployed and serving users across 180 countries.	Source checked	mongodb.com
November 30, 2025	The deployment rate has doubled from 10% to 20% in three months.	Source checked	mongodb.com
November 30, 2025	About 2.5% of the applications built in the platform's first four months were deployed: 50,000 of nearly 2 million.	Calculated	mongodb.com
July 6, 2026	The company currently sees 6 million deployments a day, half of them triggered by coding agents (Guillermo Rauch to TechCrunch).	Reported via secondary source	techcrunch.com
July 18, 2026	Of 120 repositories tagged as AI-generated, 88 (73%) were abandoned, with no activity in two to three months; a third-party study, not a vendor figure.	Reported via secondary source	xda-developers.com

DATE	AS STATED	STATUS	SOURCE
August 4, 2026	Millions of Convex instances are now running in production.	Source checked	news.convex.dev
August 12, 2026	People have created more than 60 million projects; Lovable-built apps see over 900 million visits every month; more than one-third of builders who plan to monetize are already earning revenue.	Source checked	lovable.dev
August 20, 2026	Anonymous provisioning: you have 60 minutes to build and then claim; Railway publishes no claim rate.	Source checked	railway.com
October 7, 2026	A popular developer platform managed over 300k Postgres instances on Neon with only a single engineer. Tens of thousands of projects can spin up and idle as users create apps; scale-to-zero ensures that idle databases cost you nothing while remaining instantly accessible.	Source checked	neon.com
October 7, 2026	A claimable Postgres database created for an agent's user expires after 72 hours unless it is claimed; Neon publishes no claim rate.	Source checked	neon.com
June 4, 2026	Supabase for Platforms, the product that hosts other companies' app builders, has been the fastest growing product over the last year.	Source checked	supabase.com
August 13, 2026	Exceeding \$100M revenue run-rate for Lakebase; the release does not attribute it to agents.	Source checked	databricks.com

Informs: Retained paid usage, the stock the retained-demand check (Appendix D1) models and the workflow panel's pending row (B1): read from disclosures until an operator's billing ledger matched to run ids exists. Three stages, disclosed unevenly. Creation is disclosed by several platforms and keeps rising (Supabase 60% in June 2026, 70% in October). Graduation (built to deployed) is disclosed by one, Emergent on MongoDB Atlas. Paid usage is disclosed at the vendor level, never per artifact. We had not found a platform disclosure of how many agent-created resources were still billing 30, 60 or 90 days after creation as of October 7, 2026, which is the figure the panel needs (searched October 7, 2026 across Supabase, Neon, Databricks, Convex, MongoDB, Vercel, Cloudflare, Replit, Lovable, Bolt, Railway, Render, Fly.io, PlanetScale, Turso, Upstash, Firebase, AWS and Microsoft). The platforms' own claim windows (Neon 72 hours, Railway 60 minutes) are graduation gates with no disclosed claim rates, and the only survival-style number is an independent study of AI-tagged repositories on Flathub.

Record as of October 8, 2026. Source checked means we read the primary page; reported via secondary source means we read an account of the disclosure; calculated means we computed the value from cited inputs. These labels describe how a source was checked, not independent verification of the claim. The series with comparable points are charts on the tracker's proxies section.

Appendix C — The companies: the roster, the value chain and the private record

C1 • The full research roster

The research universe is the 32 companies classified as State or Work. Other layers appear for comparison. Membership follows the published product, listing, size and data-coverage rules.

Palantir is identified separately where its size or valuation materially affects a comparison. Megaport reports half-yearly in Australian dollars and is excluded from quarterly panels that require comparable observations.

The roster includes the reason for each classification and records changes with their dates. Index membership, research coverage and holdings are different concepts. Predictions retain the universe rules under which they were registered.

The five layers of the machine-labor stack, the hyperscalers and SpaceX that span them, every public name with the reason it is where it is

L4 **Applications** 49 Where the outcome lands; the seat is the line under pressure

MAPPED, OUTSIDE THE UNIVERSE

COMPANY	LISTING	GROUP	VALUE	WHY IT IS HERE
Salesforce CRM	NYSE	Systems of record	\$188bn	The customer record and the Agentforce agents sold on top of it, priced per conversation; the largest application company and the test of the seat.
ServiceNow NOW	NYSE	Systems of record	\$139bn	The workflow record for IT, HR and customer service with agents sold as a Pro Plus tier; the clearest application-layer AI monetization.
Workday WDAY	Nasdaq	Systems of record	\$50bn	The HR and finance record with Illuminate agents priced by role; a seat business adding agent SKUs.
SAP SAP	NYSE · Xetra	Systems of record	\$243bn	The ERP record of most large enterprises, with Joule agents bundled into cloud subscriptions; the largest European software company.
Intuit INTU	Nasdaq	Systems of record	\$75bn	Accounting and tax for small business with agents inside QuickBooks; about 40% consumer, kept because the majority is business software. <i>Flag: About 40% of revenue is consumer tax; kept as majority business software.</i>

COMPANY	LISTING	GROUP	VALUE	WHY IT IS HERE
Veeva VEEV	NYSE	Systems of record	\$46bn	The commercial and clinical record of the life-sciences industry; a vertical system of record adding AI agents.
Tyler Technologies TYL	NYSE	Systems of record	\$13bn	The record of local government: courts, permits, payments; a vertical system of record.
nCino NCNO	Nasdaq	Systems of record	\$2.0bn	The lending and onboarding record of banks; a vertical system of record pricing AI by the loan.
AppFolio APPF	Nasdaq	Systems of record	\$7.2bn	Property-management record with Realm-X agents billed per unit; a vertical system of record with usage pricing.
ServiceTitan TTAN	Nasdaq	Systems of record	\$6.3bn	The operating record of the trades with AI agents sold as Pro add-ons; a vertical system of record.
Procore PCOR	NYSE	Systems of record	\$7.8bn	The construction project record priced on volume built; a vertical system of record.
Manhattan Associates MANH	Nasdaq	Systems of record	\$12bn	The supply-chain execution record; agents are being sold into warehouse and order management.
Kingdee 0268.HK	HKEX	Systems of record	\$3.3bn	Cloud ERP for Chinese enterprises with its own AI agents; the clearest listed Chinese business-software pure play.
Sage SGE.L	LSE	Systems of record	\$11bn	Accounting and payroll for small and mid-sized business with the Sage Copilot; a business-software pure play.
TechnologyOne TNE.AX	ASX	Systems of record	\$6.6bn	ERP for government, education and health in Australia and the UK; a vertical system of record moving to SaaS plus.
Xero XRO.AX	ASX	Systems of record	\$6.7bn	Small-business accounting with its JAX agent; a business-software pure play.
WiseTech Global WTC.AX	ASX	Systems of record	\$7.3bn	The logistics execution record (CargoWise) priced on transactions; a vertical system of record.
Descartes DSGX	Nasdaq · TSX	Systems of record	\$6.9bn	Logistics and customs network billed on transactions; a vertical system of record.
Kinaxis KXS.TO	TSX	Systems of record	\$3.4bn	Supply-chain planning with AI agents (Maestro); a vertical system of record.
Adobe ADBE	Nasdaq	Design and engineering	\$95bn	Creative and document software with Firefly generation priced by credits; the model is the feature and the seat is the line under pressure.

COMPANY	LISTING	GROUP	VALUE	WHY IT IS HERE
Autodesk ADSK	Nasdaq	Design and engineering	\$44bn	Design and make software with AI assistants and a move to transaction pricing; kept on the owner's call.
Figma FIG	NYSE	Design and engineering	\$10bn	Collaborative design with Figma Make generating interfaces from a description; where the application contest is most visible.
Dassault Systèmes DSY.PA	Euronext Paris	Design and engineering	\$30bn	3D design and simulation (CATIA, SOLIDWORKS, 3DEXPERIENCE); the European engineering-software pure play.
monday.com MNDY	Nasdaq	Work and communication	\$3.3bn	Work management with AI blocks billed by credits; a seat business adding usage.
Asana ASAN	NYSE	Work and communication	\$2.0bn	Work management with AI Studio billed by usage; a seat business under the most pressure from agents.
Zoom ZM	Nasdaq	Work and communication	\$27bn	Meetings, phone and contact center with an AI companion bundled; a seat business defending with AI.
RingCentral RNG	NYSE	Work and communication	\$6.4bn	Cloud phone and contact center with AI receptionist agents sold per agent; a seat business adding agent SKUs.
Five9 FIVN	Nasdaq	Work and communication	\$2.6bn	Contact-center software with AI agents billed by usage; the contact center is where agents replace seats first.
NICE NICE	Nasdaq · TASE	Work and communication	\$6.5bn	Customer-engagement software with CXone agents; the largest contact-center platform.
Box BOX	NYSE	Work and communication	\$4.6bn	Content management with Box AI agents priced per seat tier; business documents an agent reads.
DocuSign DOCU	Nasdaq	Work and communication	\$13bn	Agreements and the intelligent agreement management platform; the contract record with AI extraction.
Intapp INTA	Nasdaq	Work and communication	\$2.8bn	Deal and relationship software for professional firms with AI assistants; a vertical application.
Freshworks FRSH	Nasdaq	Work and communication	\$3.5bn	IT and customer service software with Freddy agents billed by session; a seat business adding usage.

COMPANY	LISTING	GROUP	VALUE	WHY IT IS HERE
SoundHound AI SOUN	Nasdaq	Work and communication	\$2.6bn	Voice agents for restaurants, automotive and customer service billed by usage; a pure play on agent-delivered outcomes.
Shopify SHOP	Nasdaq · TSX	Commerce and marketing	\$192bn	The commerce platform merchants and their agents sell through, paid on the volume sold; the largest usage-priced application.
Toast TOST	NYSE	Commerce and marketing	\$17bn	Restaurant operating platform paid on payment volume; a vertical application with usage pricing.
HubSpot HUBS	NYSE	Commerce and marketing	\$11bn	CRM and marketing for mid-market with Breeze agents billed by credits; a seat business adding usage.
Klaviyo KVYO	NYSE	Commerce and marketing	\$4.7bn	Marketing automation priced on contacts and messages; a usage-priced application.
Braze BRZE	Nasdaq	Commerce and marketing	\$2.8bn	Customer engagement priced on messages and data points; a usage-priced application.
Zeta Global ZETA	NYSE	Commerce and marketing	\$7.9bn	Marketing platform paid on the audiences it reaches; agents generate and place the campaigns.
Wix WIX	Nasdaq	Commerce and marketing	\$3.8bn	Website building for business with AI site generation; the application an agent can rebuild from a description.
GoDaddy GDDY	NYSE	Commerce and marketing	\$13bn	Domains, hosting and the Airo AI tools for small business; an application platform with recurring subscriptions.
Bill.com BILL	NYSE	Finance and compliance	\$4.2bn	Accounts payable and receivable for small business paid on transactions; a usage-priced finance application.
BlackLine BL	Nasdaq	Finance and compliance	\$1.6bn	Financial close software with Verity agents; a finance system of record.
Wolters Kluwer WKL.AS	Euronext Amsterdam	Finance and compliance	\$17bn	Tax, accounting, legal and health software and information; a professional-software pure play adding generative AI.
RELX REL.L	LSE · NYSE	Finance and compliance	\$56bn	Legal, risk and scientific software and analytics (Lexis+ AI); a professional-information pure play.
Varonis VRNS	Nasdaq	Finance and compliance	\$5.5bn	Data security posture for the business data agents reach, moving to SaaS; an application of the backend's governance idea to the business estate.

COMPANY	LISTING	GROUP	VALUE	WHY IT IS HERE
Appian APPN	Nasdaq	Finance and compliance	\$2.6bn	Process automation and low-code with AI agents in the process; a business-software pure play.
Samsara IOT	NYSE	Finance and compliance	\$22bn	The operating record of physical operations, priced per connected asset; a vertical system of record.
Watchlist	Kingsoft Office (STAR Market): WPS Office; consumer subscriptions are the majority, so it sits outside business software. · Amadeus (BME): Travel distribution and airline IT paid on bookings; a transaction network more than software. · Coveo (TSX): AI search and generative answering for enterprises; below the size floor. · Yelp (NYSE): Consumer reviews with a Hatch agent business for local services; consumer-majority. · Yonyou (Shanghai): Chinese ERP; A-share. · Nemetschek (Xetra): Architecture and construction software; joins if AI products are disclosed. · Temenos (SIX): Core banking software; growth slow, watch for AI pricing. · Zoho (Private): Business software suite; private.			

L3 Work 27 Executes, coordinates and accounts for machine activity

THE UNIVERSE

COMPANY	LISTING	GROUP	VALUE	WHY IT IS HERE
Cloudflare NET COMPANY CASE	NYSE	Execution and delivery	\$123bn	Executes agent code at the edge, billed by requests against a cap; the bellwether for the executing activity.
Fastly FSLY	NYSE	Execution and delivery	\$4.0bn	Edge compute and delivery billed by requests and bytes; the same meter as Cloudflare at smaller scale.
Akamai AKAM	Nasdaq	Execution and delivery	\$16bn	Edge delivery, security and a growing cloud compute line billed by usage; the slowest grower in Work, kept because agents run through its meters.
DigitalOcean DOCN	NYSE	Execution and delivery	\$16bn	Developer cloud and GPU droplets billed by the hour; where small teams run their agents.
Megaport MP1.AX	ASX	Execution and delivery	\$3.3bn	Network as a service between clouds, billed by port and by the gigabit; the only non-US backend pure play. Reports half-yearly in Australian dollars, so it sits outside the quarterly panels. <i>Flag: Half-yearly IFRS reporter in Australian dollars; in the roster and the index, outside the quarterly filings panels.</i>
Nutanix NTNX	Nasdaq	Execution and delivery	\$19bn	The private-cloud runtime agents execute on when the workload stays on premises, licensed per core; returned to Work on October 1, 2026 as an execution runtime rather than hardware.

COMPANY	LISTING	GROUP	VALUE	WHY IT IS HERE
Twilio TWLO	NYSE	Execution and delivery	\$44bn	Messages, calls and verifications billed per unit; the rail an agent uses to act in the world.
Bandwidth BAND	Nasdaq	Execution and delivery	\$2.0bn	Voice and messaging APIs billed per minute and per message; the same rail at smaller scale.
GitLab GTLB	Nasdaq	Developer platform	\$8.0bn	The pipeline where agents write, test and ship code, with agent consumption billed on top of the platform fee.
Atlassian TEAM	Nasdaq	Developer platform	\$47bn	The work record agents read and update (Jira, Confluence, Bitbucket) and the Rovo agents that run against it.
JBFrog FROG	Nasdaq	Developer platform	\$12bn	The artifact registry every build and model passes through, billed by storage and transfer.
UiPath PATH	NYSE	Developer platform	\$6.8bn	Agents, robots and orchestration metered per run, decision and token; moved from the application layer on September 24, 2026.
Datadog DDOG COMPANY CASE	Nasdaq	Observability	\$98bn	Observes the machine hour on hosts, events and model calls; the bellwether for the observing activity.
Dynatrace DT	NYSE	Observability	\$17bn	Observability and automatic remediation billed by consumption; the audit trail of production systems.
PagerDuty PD	NYSE	Observability	\$1.4bn	Incident response and the automation that runs when something breaks; where agent failures are escalated.
Okta OKTA	Nasdaq	Identity and access	\$35bn	Identity for people and now for agents, billed per identity; every agent action starts with a credential.
SailPoint SAIL	Nasdaq	Identity and access	\$12bn	Governs which identities, human and machine, may touch which systems; agent identity is its fastest-growing line.
Zscaler ZS	Nasdaq	Identity and access	\$32bn	Zero-trust access and AI traffic controls billed by the traffic they inspect; the gate an agent passes to reach a system.
Netskope NTSK	Nasdaq	Identity and access	\$7.0bn	Secure access and data controls for cloud and AI traffic; the same gate from the data side.

COMPANY	LISTING	GROUP	VALUE	WHY IT IS HERE
CrowdStrike CRWD	Nasdaq	Security of the runtime	\$270bn	Protects the endpoints and workloads agents run on, billed per sensor and by data; agent and cloud security modules drive growth.
SentinelOne S	NYSE	Security of the runtime	\$8.0bn	Autonomous endpoint and cloud protection; the same job as CrowdStrike at smaller scale.
Tenable TENB	Nasdaq	Security of the runtime	\$4.0bn	Exposure management across the systems agents touch, billed per asset.
Qualys QLYS	Nasdaq	Security of the runtime	\$6.6bn	Vulnerability and compliance scanning billed per asset; the audit of what the runtime is exposed to.
Palo Alto Networks PANW	Nasdaq	Security of the runtime	\$324bn	Network, cloud and AI runtime security; the largest security platform and the one that bills AI protection as its own line.
Rubrik RBRK	NYSE	Data protection and governance	\$24bn	Backup, recovery and governance of other systems' data, billed by the data protected; the record an agent cannot destroy.
Commvault CVLT	Nasdaq	Data protection and governance	\$6.2bn	Backup and recovery across clouds, increasingly billed by consumption; the same job as Rubrik with an older base.
AvePoint AVPT	Nasdaq	Data protection and governance	\$3.0bn	Governance, backup and migration of the Microsoft estate agents work in, billed by the data under management.

Watchlist

Check Point (Nasdaq): Network security with slow growth and an appliance base; moved to watch in the October 1, 2026 cut. · **Sinch** (Nasdaq Stockholm): Communications APIs like Twilio; AI revenue undisclosed and growth flat. · **Wangsu** (Shenzhen): Chinese CDN and edge; A-share, dropped October 1, 2026. · **OVHcloud** (Euronext Paris): European cloud; AI revenue undisclosed, dropped October 1, 2026. · **UCloud** (STAR Market): Chinese neutral cloud; no data on the feed. · **Cyber Security Cloud** (TSE): Web application security billed by usage; below the size floor. · **BlackBerry** (NYSE · TSX): QNX is the embedded runtime of vehicles and devices; the rest of the company is not backend.

L2 State 5 Keeps information: the record, the memory, the data an agent reasons over

THE UNIVERSE

COMPANY	LISTING	GROUP	VALUE	WHY IT IS HERE
Snowflake SNOW COMPANY CASE	NYSE	State	\$118bn	Governed data an agent reasons over, billed by warehouse time and AI credits; the bellwether for the reasoning activity.

COMPANY	LISTING	GROUP	VALUE	WHY IT IS HERE
MongoDB MDB COMPANY CASE	Nasdaq	State	\$28bn	The operational database agents read and write, billed by cluster capacity; the bellwether for the remembering activity.
Elastic ESTC	NYSE	State	\$9.5bn	Search and vector retrieval over logs and documents, billed by the resources the cluster uses; agent memory and retrieval land here.
Backblaze BLZE	Nasdaq	State	\$0.8bn	Object storage billed per byte stored and served; the smallest name in the universe, kept as the pure storage meter.
Palantir PLTR	Nasdaq	State	\$430bn	The platform where enterprise data is modeled and agents act on it; moved from the application layer because it sells the record, not an outcome. Contracts are fixed-price, so it is flagged for every mean. <i>Flag: Fixed-price contracts and a market value that distorts any mean; medians are quoted.</i>

L1 Intelligence 2 Tokens and inference; priced, competitive, deflating MAPPED, OUTSIDE THE UNIVERSE

COMPANY	LISTING	GROUP	VALUE	WHY IT IS HERE
Z.ai 2513.HK	HKEX	Frontier models	\$36bn	Trains and serves the GLM models and sells inference by the token and by enterprise license; one of two listed pure plays on the model layer.
MiniMax 0100.HK	HKEX	Frontier models	\$10bn	Trains and serves its own models and sells inference by the token to developers and consumer products; the other listed pure play.

Watchlist	Anthropic (Private): Frontier lab; private. · Moonshot AI (Private): Kimi models; private. · StepFun (Private): Step models; private. · DeepSeek (Private): Open-weight frontier models; private. · SenseTime (HKEX): Listed, but generative AI is a share of a wider vision and smart-city business; joins if the model business becomes the majority.			
-----------	--	--	--	--

L0 Computation 25 Compute, memory, networking, systems, power and rented capacity MAPPED, OUTSIDE THE UNIVERSE

COMPANY	LISTING	GROUP	VALUE	WHY IT IS HERE
Nvidia NVDA	Nasdaq	Silicon	\$5532bn	Accelerators, networking and the CUDA software the machine hour runs on; data-center revenue is most of the company and has grown sequentially for fourteen straight quarters.

COMPANY	LISTING	GROUP	VALUE	WHY IT IS HERE
Broadcom AVGO	Nasdaq	Silicon	\$1671bn	Custom AI accelerators for the hyperscalers and the Ethernet switching between them; AI semiconductor revenue is the growth of the company.
Marvell MRVL	Nasdaq	Silicon	\$231bn	Custom silicon and electro-optics for AI data centers; the data-center segment is now most of revenue and the rest of the company is flat.
Cerebras CBRS	Nasdaq	Silicon	\$42bn	Wafer-scale inference and training systems sold as hardware and as a cloud; a pure play on inference compute with one large customer.
AMD AMD	Nasdaq	Silicon	\$998bn	Instinct accelerators and the EPYC servers that host them; admitted as a scale leader, since gaming and client still dilute the AI share.
TSMC TSM	NYSE · TWSE	Silicon	\$2366bn	Fabricates nearly every leading accelerator; high-performance computing is over half of revenue. Admitted as a scale leader.
Micron MU	Nasdaq	Memory, test and packaging	\$1203bn	High-bandwidth memory for accelerators; HBM moved from nothing to a multi-billion-dollar run rate in two years and sets the growth of the company.
Teradyne TER	Nasdaq	Memory, test and packaging	\$63bn	Test systems for AI accelerators, HBM and the custom silicon behind them; compute test is the growth of the semiconductor segment.
Unimicron 3037.TW	TWSE	Memory, test and packaging	\$60bn	ABF substrates for accelerators and switch silicon; the AI substrate is the segment that is growing.
SK hynix 000660.KS	KRX	Memory, test and packaging	\$956bn	The leading HBM supplier, with most of the HBM3E market; admitted as a scale leader because commodity DRAM and NAND remain large.
ASML ASML	Nasdaq · Euronext	Memory, test and packaging	\$698bn	The only EUV lithography supplier, the tool every leading accelerator and HBM wafer passes through. Admitted as a scale leader.
Credo CRDO	Nasdaq	Networking and optics	\$36bn	Active electrical cables and SerDes that connect accelerators inside the rack; revenue roughly tripled on AI cluster builds.

COMPANY	LISTING	GROUP	VALUE	WHY IT IS HERE
Astera Labs ALAB	Nasdaq	Networking and optics	\$61bn	PCIe, CXL and Ethernet connectivity silicon between accelerators, CPUs and memory; the whole company is the AI rack.
Coherent COHR	NYSE	Networking and optics	\$56bn	Optical transceivers and the indium phosphide lasers inside them; data-center optics are most of the growth.
Lumentum LITE	Nasdaq	Networking and optics	\$76bn	Lasers and transceivers for AI data-center interconnect; cloud and networking revenue has displaced the telecom business.
Super Micro Computer SMCI	Nasdaq	Systems and ODM	\$27bn	Rack-scale accelerator servers and liquid cooling; AI platforms are about two thirds of revenue.
Celestica CLS	NYSE - TSX	Systems and ODM	\$42bn	Data-center switches and servers for the hyperscalers; the connectivity and cloud segment is most of revenue and all of the growth.
Quanta Computer 2382.TW	TWSE	Systems and ODM	\$40bn	The largest AI server ODM; cloud servers passed notebooks as its largest business.
Wiwynn 6669.TW	TWSE	Systems and ODM	\$37bn	Hyperscale servers and racks only, built to order for the cloud providers; a pure play on the data-center build.
Wistron 3231.TW	TWSE	Systems and ODM	\$19bn	Assembles accelerator baseboards and AI servers for Nvidia and the hyperscalers; AI servers drive the growth.
Dell Technologies DELL	NYSE	Systems and ODM	\$358bn	AI-optimized servers with a multi-billion-dollar backlog; admitted as a scale leader, since PCs and storage are still most of revenue.
Vertiv VRT	NYSE	Power and cooling	\$93bn	Power, thermal and liquid cooling for AI data centers; orders and backlog track the accelerator build.
Asia Vital Components 3017.TW	TWSE	Power and cooling	\$43bn	Liquid cooling and thermal modules for accelerator racks; data-center cooling is the growth of the company.
CoreWeave CRWV	Nasdaq	Capacity	\$48bn	GPU cloud sold by the hour under multi-year contracts; a pure play on rented accelerator capacity.
Nebius NBIS	Nasdaq	Capacity	\$57bn	GPU cloud and the data centers behind it; AI infrastructure is nearly all of revenue.

COMPANY	LISTING	GROUP	VALUE	WHY IT IS HERE
Watchlist				Iluvatar CoreX (HKEX): GPU designer listed in Hong Kong; price only on the data feed, no income statements yet. · Biren Technology (HKEX): GPU designer listed in Hong Kong; price only on the data feed, no income statements yet. · Moore Threads (STAR Market): GPU designer on the Shanghai STAR Market; no price or statements on the data feed. · MetaX (STAR Market): GPU designer on the Shanghai STAR Market; no price or statements on the data feed. · Enflame (STAR Market): Accelerator designer on the Shanghai STAR Market; no price or statements on the data feed. · TeraWulf (Nasdaq): Bitcoin miner converting its power to AI hosting; joins when hosting is the majority of revenue. · Isu Petasys (KRX): High-layer circuit boards for AI switches and servers; joins when the AI share is disclosed. · Hanmi Semiconductor (KRX): Thermal-compression bonders for HBM; revenue is the HBM cycle, cut on size and concentration. · Nscale (Private): European GPU cloud; private.

CROSS-LAYER **Hyperscalers and SpaceX** 8

Span several layers; the hyperscalers measured on their cloud segments only, SpaceX unmeasured until it reports an AI segment

REFERENCE GROUP, OUTSIDE THE INDICES AND MEDIANS				
COMPANY	LISTING	CLOUD SHARE	VALUE	WHY IT IS HERE
Microsoft MSFT	Nasdaq	Intelligent Cloud about 42% of revenue	\$3809bn	Azure spans Computation to Work and hosts OpenAI; measured on Intelligent Cloud, about 42% of fiscal-2026 revenue.
Oracle ORCL	NYSE	Cloud infrastructure about 27%; all cloud about 50%	\$396bn	OCI rents accelerators and runs the database; measured on cloud infrastructure, about 27% of revenue, with all cloud about half.
Baidu BIDU	Nasdaq · HKEX	AI Cloud about 23% of revenue	\$30bn	The AI Cloud segment sells compute and the ERNIE models; about 23% of revenue.
Amazon AMZN	Nasdaq	AWS about 18% of revenue	\$2680bn	AWS spans every layer from chips to agent memory; about 18% of revenue and most of the operating income.
Alibaba BABA	NYSE · HKEX	Cloud Intelligence about 15% of revenue	\$258bn	Cloud Intelligence sells compute and the Qwen models; about 15% of revenue.
Alphabet GOOGL	Nasdaq	Google Cloud about 15% of revenue	\$4164bn	Google Cloud spans TPUs to Gemini to agent memory; about 15% of revenue.

COMPANY	LISTING	CLOUD SHARE	VALUE	WHY IT IS HERE
Tencent 0700.HK	HKEX	Not disclosed; inside a segment of about 31%	\$495bn	Tencent Cloud sits inside a fintech and business services segment of about 31% of revenue and is not disclosed on its own; carried without a cloud measure.
SpaceX SPCX	Nasdaq	No AI segment disclosed	\$1971br	Listed June 12, 2026. Spans Intelligence, Applications and Computation through xAI (acquired February 2026), Cursor (acquired August 14, 2026) and its data-centre build; launch and Starlink are most of revenue and no AI segment is disclosed, so it sits outside every layer line and the universe until it reports one.

108 public pure plays across the five layers, of which the 32 in State and Work are the universe the indices, medians and bets run on; 7 hyperscalers and SpaceX as the cross-layer reference group; 29 names on the watchlists. Market values are the latest close available on October 1, 2026, Financial Modeling Prep quotes converted to US dollars at the same day's rates. Rules: Listed anywhere, and present with a price and income statements on the site's market-data feed. A pure play on the layer itself: the layer's product is the majority of revenue and the source of growth. No micro-caps: about US\$1 billion for the backend, about US\$2 billion for Applications. Applications is business software only; consumer, fintech, healthcare delivery and ad-tech stay out. Computation needs the AI build to be the majority of revenue and the growth already reported; five scale leaders are admitted by size and marked. Mainland A-shares are not held; they go to the watchlists. Roster as of October 1, 2026.

C2 · From the value of the work to backend revenue

The economic value of an agent's work begins with accepted output, including time saved and tasks newly made affordable. Review and integration consume part of the benefit. Customers retain part, and the remainder is divided among providers across the system.

The knowledge-work wage pool is therefore a scale comparison. It cannot be converted directly into backend revenue. Application revenue and the infrastructure purchased to deliver it must not be counted twice in an aggregate market total.

Accepted value	Customer share	Stack allocation
saved cost + new output	value retained by the buyer	spending across providers
existing work → newly economic tasks	review and integration → required return on spend	models and apps → compute, state and work

Backend revenue

the portion retained by this market

A wage-pool sensitivity is not an independently measured software TAM.

C3 · The private universe: the dated record

The private-company table records the latest sourced financing, operating disclosure or relevant launch for each company. Round prices, primary capital, secondary transactions and debt are identified separately.

Companies that leave the active map remain in the record with the reason. A financing or acquisition is evidence of a transaction. Paid use and retention are the evidence for the demand thesis.

Runtime

COMPANY	LATEST DATED EVENT
Temporal	September 14, 2026 Series E led by Lightspeed, with Wellington Management participating; more than 4,300 paying customers; 1.9 trillion billable actions in August 2026; run-rate up more than 200% year over year (GeekWire reports the run-rate passed \$250M). \$550M at \$12.55B source
Vercel	June 17, 2026 Ship 2026 release: coding agents now trigger more than half of deployments to Vercel infrastructure, from fewer than 3% six months earlier; AI Gateway carries roughly 2 to 20 trillion tokens a month. >50% of deployments by agents source
Modal	May 21, 2026 Series C led by Redpoint Ventures and General Catalyst; annualized revenue about \$300M, up from \$60M in September 2025 (Reuters exclusive). \$355M at \$4.65B source
E2B	September 15, 2026 Company site displays a cumulative usage counter of 1B+ sandboxes started (observed on the updates page on 2026-09-15; no dated post announces the milestone). 1B+ sandboxes started source
Browserbase	June 17, 2025 Series B led by Notable Capital with CRV and Kleiner Perkins; Director, a natural-language web-automation tool, launched the same day. \$40M at \$300M post-money source
LangChain	October 21, 2025 Series B led by IVP; LangChain and LangGraph 1.0 shipped the same month. \$125M at \$1.25B source
Replit	March 11, 2026 Series D led by Georgian; Replit Agent 4 launched alongside; company targets \$1B ARR by the end of 2026. \$400M at \$9B source
Lovable	August 12, 2026 Series C led by Menlo Ventures, co-led by EQT's Scaleup Europe Fund; 60 million projects since the November 2024 launch; 900 million monthly visits to Lovable-built apps. \$400M at \$13.3B source
Cognition	September 25, 2026 Annualized revenue run rate passes \$1B, from \$492M in May 2026; GE Aerospace, Rivian, Rohlik and Exa named as customers. \$1B run-rate revenue source
n8n	August 5, 2026 n8n Agents in preview on n8n Cloud: agents defined once and run through chat, as a workflow node, in Slack or on a schedule. source

Inference cloud

COMPANY	LATEST DATED EVENT
Fireworks AI	July 16, 2026 Series D led by Atreides, Index and TCV; annualized revenue run rate \$1B; more than 40 trillion tokens served a day. \$1.505B at \$17.5B source

COMPANY	LATEST DATED EVENT
Baseten	June 22, 2026 Series F led by Altimeter, Conviction and Spark; revenue up 20x over the past year. \$1.5B at \$13B source
Together AI	July 1, 2026 Series C led by Aramco Ventures; new investors commit more than 500 MW of compute capacity; TechCrunch puts the post-money at \$8.3B and annual bookings above \$1.15B. \$800M at \$8.3B (valuation per TechCrunch) source
fal	May 19, 2026 Names AWS its preferred cloud; serves more than 2.5 million developers with more than 1,000 production image, video, audio and 3D models through one API. 2.5M developers source

Control plane

COMPANY	LATEST DATED EVENT
Grafana Labs	February 13, 2026 The Information reports a GIC-led round at a \$9B valuation, up from \$6.6B (aggregators later list a \$250M Series E closing 2026-03-12; no company release found). \$9B valuation (reported) source · reported
Vanta	April 29, 2026 ARR passes \$300M nine months after \$200M; 16,000 customers; growth rate rose in each of the prior four quarters. \$300M ARR; 16,000 customers source
Braintrust	February 17, 2026 Series B led by ICONIQ; customers include Notion, Replit, Cloudflare, Ramp and Dropbox (valuation about \$800M per Axios). \$80M at about \$800M source
WorkOS	March 2, 2026 Series C led by Meritech and Sapphire; named customers include OpenAI, Anthropic, xAI, Cursor, Vercel, Replit, Baseten and Temporal. \$100M at \$2B source
PostHog	July 31, 2026 Replay Vision: an AI layer scans session recordings for frustration and hands findings to a background agent that opens pull requests; two such pull requests merged. source
Clerk	March 3, 2026 Ships agent skills and an MCP server and redesigns its core auth APIs to work for coding agents as well as human developers. source

State

COMPANY	LATEST DATED EVENT
Databricks	August 13, 2026 Round closes; revenue run-rate \$7B growing more than 80% year over year; Lakebase above \$100M run-rate; more than 1,000 customers above \$1M. \$5B at \$190B source
ClickHouse	May 27, 2026 Annualized revenue \$250M, tripled year over year; more than 4,000 customers including Anthropic and Meta. \$250M annualized revenue source
Redis	October 9, 2025 Acquires Featureform to deliver structured, real-time context to agents; price undisclosed. source
Supabase	October 2, 2026 Supabase announced \$150M in funding and its acquisition of Turso. The round includes employee liquidity; the split between new company capital and secondary transactions is not disclosed. Acquisition terms are not disclosed. \$150M announced; primary/secondary split undisclosed source
Convex	August 4, 2026 Series B led by Insight Partners; total raised \$110.5M; nearly 2 million applications from about 500,000 developers; customers include OpenAI and Zapier; release claims 90% of AI-built apps on a traditional database corrupted data versus 0% on Convex (vendor claim). \$57M Series B source

COMPANY	LATEST DATED EVENT
Pinecone	August 6, 2026 Nexus knowledge engine for agents reaches general availability; more than 10,000 customers and 1 million developers; company reports a 47.4% pass rate on Sierra's tau-Knowledge benchmark versus 46.4% for GPT-5.5 alone. 10,000+ customers; 1M developers source
Neo4j	February 3, 2026 Aura Agent reaches general availability, free for February 2026. source
Spacetime	May 14, 2026 Revised benchmark post reports 303,920 +/- 4,712 transactions per second on a contended bank-transfer workload (alpha 1.5) on one Intel 14900K node, against 3,188 for Node.js plus SQLite, 961 for Node.js plus Postgres and 127 for Convex; vendor cost table puts Spacetime at about \$3.60 a month per 1,000 TPS. 303,920 TPS contended (vendor benchmark) source

C4 · State and Work: the economic questions

State keeps data, memory and records that can remain after a task ends. Work runs software and manages the activity around it. Both can be priced by capacity, usage or subscription, depending on the product.

The questions are practical: what persists, what the customer pays for, and what could make that spending move elsewhere. Compression and portability can reduce data costs. Sampling and bundling can reduce monitoring revenue. The model tests those effects separately from growth in hours.

ECONOMIC QUESTION	STATE	WORK
What persists?	Data, memory, context and permissions	Deployments, execution and monitoring
What gets billed?	Compute, storage and product-specific units	Hosts, ingestion, requests and subscriptions
What limits reach?	Compression and open alternatives	Sampling, bundling and control-plane concentration

Appendix D — The record: downside, registered calls and disclosures

D1 • What would change our view: sensitivities, risks and evidence to watch

Adoption and active time drive hours demanded. Compute can limit delivery. Provider share and pricing determine agent revenue. Costs and dilution determine how much revenue becomes value for owners.

The main commercial risk is that machine work grows while the selected providers earn less. The sensitivity table changes the hourly bill, provider-use share and price separately. The model also tests slower adoption, limited provider revenue and a capacity path that stops expanding after 2029.

Software built by agents can keep generating infrastructure bills after the task ends. We estimate this continuing demand separately to check the model. The published scenarios already include it in existing-business revenue, so we do not add it again.

If activity rises but paid use does not, reduce the assumed share of work using these providers and the revenue earned per hour. If repeated use is weak, reduce adoption or active hours. If applications are abandoned, reduce the spending that continues after they are built.

Five inputs moved alone: what each does to the four companies’ combined 2032 revenue

INPUT MOVED ALONE	AGENTIC WORKFORCE 2032	MACHINE ECONOMY 2032
Published path	\$71bn	\$124bn
Cloudflare’s hourly bill at 5 cents an hour instead of 1	\$87bn +16	\$175bn +52
Datadog’s hourly bill at a fully traced hour, 35 cents instead of 5	\$90bn +20	\$201bn +77
Provider-use shares held at their 2025 assumptions	\$67bn –4	\$114bn –9
Reach halved from our path	\$64bn –6	\$103bn –21
Price at 30% of today’s at ten times the volume	\$57bn –14	\$74bn –50

Combined 2032 revenue of the four company cases, from the October 6, 2026 model. Each row changes one assumption and holds the others fixed. These are scenario results.

Discounted terminal enterprise value under selected 2032 revenue multiples

SCENARIO · 2032 REVENUE	EXIT AT 5X	EXIT AT 6.6X	EXIT AT 8X
Agentic Workforce · \$71bn	\$200bn 0.6× the September 8, 2026 reference	\$263bn 0.8× the September 8, 2026 reference	\$319bn 1.0× the September 8, 2026 reference
Machine Economy · \$124bn	\$349bn 1.1× the September 8, 2026 reference	\$460bn 1.5× the September 8, 2026 reference	\$558bn 1.8× the September 8, 2026 reference
Multiple matching the dated reference after discounting	Agentic Workforce about 7.8x · Machine Economy about 4.5x		

Discounted terminal enterprise value: each scenario's 2032 combined revenue times the exit multiple, discounted six years at 10%, against \$313 billion of combined enterprise value on September 8, 2026. 5x was the backend's forward multiple at the June 30, 2026 quarter end and 6.6x at September 30; 8x sits below the forward multiples of three of the four. This includes only the terminal enterprise value, not cash flows earned along the way. It is not an equity return or price target; debt, cash, dilution and distributions would also matter. The multiples are selected assumptions, not predicted exit prices. Modeled.

Spending on applications that stay running after an agent's task ends. The 30 cents per dollar of agent work and 40% annual churn are assumptions.

Continuing spending at the published assumptions (30¢ of yearly run-rate per \$1 of agent work, 40% annual loss rate), included in existing-business revenue

YEAR	AGENT REVENUE, AW / ME (\$BN)	CONTINUING ANNUAL SPENDING, AW / ME (\$BN)	SHARE OF EXISTING-BUSINESS GROWTH SINCE 2025, AW / ME
2027	0.11 / 0.12	0.01 / 0.01	0.1% / 0.1%
2028	0.31 / 0.45	0.04 / 0.04	0.2% / 0.2%
2029	0.86 / 1.66	0.12 / 0.16	0.4% / 0.5%
2030	2.30 / 5.72	0.33 / 0.59	0.9% / 1.2%
2031	6.00 / 18.17	0.89 / 2.07	2.2% / 3.6%
2032	15.03 / 53.37	2.33 / 6.69	5.4% / 12%

Four assumptions for continuing spending, and the effect if some were additional revenue

CASE	WHAT BOUNDS IT	2032 CONTINUING ANNUAL SPENDING, AW / ME (\$BN)	SHARE OF EXISTING- BUSINESS GROWTH	2032 TOTAL IF HALF IS ADDITIONAL	2032 TOTAL IF ALL IS ADDITIONAL
Low	15¢ of yearly run-rate per \$1 of agent work, half churns a year: Replit hosted 1 in 20 agent-built apps; half of agent-created Lakebase databases live under 10 seconds	1.1 / 3.2	2.6% / 5.6%	\$71.3bn / \$125.2bn +0.8% / +1.3%	\$71.8bn / \$126.8bn +1.6% / +2.6%

CASE	WHAT BOUNDS IT	2032 CONTINUING ANNUAL SPENDING, AW / ME (\$BN)	SHARE OF EXISTING- BUSINESS GROWTH	2032 TOTAL IF HALF IS ADDITIONAL	2032 TOTAL IF ALL IS ADDITIONAL
Mid, the preset	30¢ per \$1 and 40% annual churn are assumptions, not measured retention. Agent-created deployments and databases measure creation, not continued paid use.	2.3 / 6.7	5.4% / 12%	\$71.9bn / \$126.9bn +1.6% / +2.7%	\$73.0bn / \$130.2bn +3.3% / +5.4%
High	60¢ per \$1, 30% churn: enterprise pipelines and clusters that stay; Snowflake's Cortex Code accounts consuming more of the core	4.9 / 13.9	11% / 24%	\$73.2bn / \$130.5bn +3.5% / +5.6%	\$75.6bn / \$137.5bn +6.9% / +11.3%
Stress	\$1 per \$1, 30% churn	8.2 / 23.2	19% / 40%	\$74.8bn / \$135.1bn +5.8% / +9.4%	\$78.9bn / \$146.7bn +11.6% / +18.8%

Continuing demand is annual spending on the software agents leave running. The published scenarios include it in existing-business revenue, so it is not added again. The table tests alternative assumptions and shows the effect if some spending were additional. Creation and deployment disclosures help frame the question, but do not measure continued paid use. The 30-cent spending assumption and 40% annual loss rate are model judgments, not observed retention.

Open question, the evidence that would settle it, and what it would do to the framework

OPEN QUESTION	EVIDENCE TO WATCH	IMPLICATION FOR THE FRAMEWORK
Do agent-created databases and applications survive into paid production?	Vercel's two measures, kept separate: 30% of the applications on the platform originated from agents (April 13, 2026), and coding agents trigger more than half of deployments (June 17, 2026, from under 3% six months earlier); neither measures retained use; paid conversion of agent-created Supabase and Neon databases; MongoDB's Atlas ARR from customers with an AI use case (40%).	Weak conversion cuts the link between infrastructure creation and durable demand: the retained-demand term stays a check, and the reach shares rise more slowly.
Does agent activity become paid consumption at the four, or stay inside headroom and free tiers?	Product revenue growth against Cortex Code accounts (Snowflake), Workers developers (Cloudflare), AI-native customers (MongoDB, Datadog); commitments and AI cohorts each quarter.	More activity with limited paid demand lowers assumed provider use and revenue per hour; the countercase lands at \$39.7 billion in 2029 and \$58.9 billion in 2032.

OPEN QUESTION

EVIDENCE TO WATCH

IMPLICATION FOR THE FRAMEWORK

How fast does the unit price fall as volume multiplies?

Disclosed price per unit at the meters: Datadog indexed logs, MongoDB revenue per query, Snowflake revenue per query; AI Credit and inference prices.

A fall of 70% per tenfold instead of 35% to 50% cuts 2032 revenue by about 40% in the Machine Economy, \$123.5 billion to \$73.9 billion (D1).

Are agents permitted, integrated and used repeatedly, not just capable?

Repeat usage and retention of agent products; human intervention per completed task (Agentforce escalations, 32%); the share of US employees using AI daily (Gallup) and of firms using AI (Census), the nearest published proxies for agent use.

Lower autonomy shortens active duration: hours per agent fall and the crossover moves later; the constrained case misses it within the period.

Does one hour of work burn fewer tokens every year, and does that cut the bill or just the cost?

Our monthly workload trace; cache shares at the gateways; tokens per request on OpenRouter and in Datadog's traces.

Fewer tokens an hour raise the hours a given build can serve (the compute ceiling) and lower the token forecast; the receipt is per touched hour, so revenue holds unless the hours fall.

Does the core business hold its reported growth through 2028?

The four's quarterly growth against the 36%, 30%, 35%, 36% reported in mid-2026; guidance for the next quarter; the AI-native cohorts.

Two of the four decelerating for two consecutive quarters fails the near-term claim; nine tenths of the 2029 gap to consensus is this assumption.

A risk is answered by a test, not by an argument: each row names what would settle it and what the model does if it goes the other way. The first two rows test durable paid demand; the fourth decides the headline. Re-graded each quarter with the bets.

Demand disappoints

- Too little accepted customer value
- High review and failure costs
- Pilots fail to become recurring work

Reach disappoints

- Local, open or bundled alternatives
- A smaller receipt and a steeper price path
- Fewer hours reach the four

Returns disappoint

- Serving costs offset sales growth
- Reinvestment and dilution absorb cash
- Valuation already anticipates success

FAILURE SIGNAL	REVIEW WINDOW	ACTION
Accepted task cost fails to improve	Two matched quarterly operating-panel observations	Cut adoption and hours per agent; find where the review or failure cost sits.
Autonomy does not spread	Two consecutive quarters of the readings under size	Cut agents per user and the round-the-clock share; the Machine Economy loses its path first.
The receipt shrinks	Two matched quarters of disclosed volume and revenue per unit	Steepen the price path and cut the receipt.
Backend growth advantage disappears	Two comparable reported quarters	Reassess the relative infrastructure-demand argument.
Cash conversion deteriorates	Two trailing four-quarter observations	Reassess serving costs, reinvestment and diluted shareholder economics.
The workload migrates	A documented annual operating-panel review	Cut the share of hours that reach the four; category necessity is not incumbent durability. Agent-native runtimes that ship their own metrics, tracing and audit log count here: an application built on one may never send a span to a separate observability vendor.

D2 · The launch predictions and their grading rules

These are the original launch bets, registered October 8, 2026. Each has a stated measure, observation period and grading deadline. The full rule appears beside each bet in the tracker.

H1 asks whether active machine duration passes the fixed 2026 paid-human-work benchmark by the end of 2032. It does not equate machine and human output. A robust hit requires the lower sensitivity bounds of both the provider-by-provider and OpenRouter-based estimates above 7.5 trillion annual hours; a miss requires both upper bounds below 6.5 trillion. Overlap or disagreement is unresolved. Initial grade January 31, 2033; one final restatement of the last complete week of December 2032 by June 30, 2033 may remain unresolved. Independent method review is due September 30, 2032.

B15 applies the same uncertainty principle to 100 million hours a day in the last complete week of December 2027: initial grade January 31, 2028, final grade June 30, 2028. Its method review is due September 30, 2027. B2 observes the June 2027 quarter and allows filings through September 30,

2027. Missing or incomparable filings leave these grades unresolved.

B3 sets an absolute target of 17 companies using one inclusion rule. No comparable launch baseline is asserted. The model scenarios, the headline crossover and company-revenue outcomes are separate claims: success in one does not establish the others. The 65/35 weights are conditional on the two published adoption scenarios, not a probability that the crossover happens.

ID	PREDICTION	GRADE BY	STATUS
H1	Machine hours exceed all paid human work, 7.0 trillion hours a year, by the 2032 reading.	2033-01-31	OPEN
B1	The backend index beats the Computation layer over the 12 months from October 1, 2026.	2027-10-01	OPEN
B2	Median backend revenue growth improves at least three points from the quarter ended June 30, 2026.	2027-09-30	OPEN
B3	At least 17 of the 32 backend companies disclose an aggregate AI-attributable revenue, ARR, run-rate or bookings amount, or a share of revenue.	2027-09-30	OPEN
B4	A publicly evaluated model reaches an 80% task horizon of at least eight human-expert hours.	2027-06-30	OPEN
B5	OpenRouter's daily tokens triple.	2027-09-08	OPEN
B7	AI agents performing user-directed work exceed human-originated requests in the disclosed traffic of at least two major platforms.	2031-12-31	OPEN
B8	The 32-name backend reports more than \$250 billion of annual recognized revenue.	2032-03-31	OPEN
B11	The four bellwethers' combined revenue growth stays at or above 30% year over year in each of the next four reported quarters.	2027-09-30	OPEN

ID	PREDICTION	GRADE BY	STATUS
B12	Cloudflare, MongoDB, Snowflake and Datadog report more than \$40 billion of combined revenue for the 2029-aligned reporting periods, against the registered \$32.7 billion consensus snapshot.	2030-03-31	OPEN
B14	The backend is the best-performing layer of the stack over the three years from October 1, 2026.	2029-10-01	OPEN
B15	By the end of 2027, AI agents perform more than 100 million active hours a day worldwide.	2028-01-31	OPEN
DEADLINE	LONG CALL		
2036-12-31	A usage-oriented infrastructure business exceeds \$100B of annual recognized revenue.		
2038-12-31	The public backend universe reaches \$1T of annual recognized revenue.		
2040-12-31	An external index provider launches an agent-infrastructure index.		

B3: one disclosure rule, every company visible

One rule for the 32-name launch roster: an aggregate AI revenue, ARR, run-rate or bookings amount, or a share of revenue. AI-native customer cohorts are labeled separately. Growth rates, shares of growth, share of bookings, customer counts and single contracts do not qualify.

This is an evidence register, not a complete launch baseline. Four qualifying primary records were checked for this release. Other rows are explicitly unverified or show why a particular record is excluded. The bet is the absolute target of 17 by its observation deadline.

COMPANY	RECORD STATUS	DATED EVIDENCE	REASON
Snowflake (SNOW)	This record excluded	2026-09-02: about half of the 7-point product revenue growth acceleration (30% to 37%) attributed to AI products; implied about 3.5 points (derived) Source	Share of growth is excluded. An earlier absolute amount or revenue share may qualify after primary-source review.
MongoDB (MDB)	No qualifying record established	No qualifying primary record established here.	The existing record contains adoption, qualitative commentary or no amount. This is not proof that all disclosures lack a qualifying figure.

COMPANY	RECORD STATUS	DATED EVIDENCE	REASON
Elastic (ESTC)	No qualifying record established	No qualifying primary record established here.	The existing record contains adoption, qualitative commentary or no amount. This is not proof that all disclosures lack a qualifying figure.
Backblaze (BLZE)	This record excluded	2026-05-04: more than one third of all new bookings from AI Source	Share of bookings is not a bookings amount or share of revenue.
Palantir (PLTR)	Unverified	No qualifying primary record established here.	No complete primary-source four-quarter review is recorded under this rule.
Cloudflare (NET)	No qualifying record established	No qualifying primary record established here.	The existing record contains adoption, qualitative commentary or no amount. This is not proof that all disclosures lack a qualifying figure.
Fastly (FSLY)	No qualifying record established	No qualifying primary record established here.	The existing record contains adoption, qualitative commentary or no amount. This is not proof that all disclosures lack a qualifying figure.
Akamai (AKAM)	No qualifying record established	No qualifying primary record established here.	The existing record contains adoption, qualitative commentary or no amount. This is not proof that all disclosures lack a qualifying figure.
DigitalOcean (DOCN)	Qualifying record verified	2026-08-04: \$234 million AI Customer ARR Source	Customer cohort; includes infrastructure that may not itself be an AI workload.
Megaport (MP1.AX)	Unverified	No qualifying primary record established here.	No complete primary-source four-quarter review is recorded under this rule.
Nutanix (NTNX)	Unverified	No qualifying primary record established here.	No complete primary-source four-quarter review is recorded under this rule.
Twilio (TWLO)	No qualifying record established	No qualifying primary record established here.	The existing record contains adoption, qualitative commentary or no amount. This is not proof that all disclosures lack a qualifying figure.

COMPANY	RECORD STATUS	DATED EVIDENCE	REASON
Bandwidth (BAND)	No qualifying record established	No qualifying primary record established here.	The existing record contains adoption, qualitative commentary or no amount. This is not proof that all disclosures lack a qualifying figure.
GitLab (GTLB)	This record excluded	2026-09-01: Duo Agent Platform paid consumption run rate up about 50% quarter-over-quarter; platform-wide paid consumption run rate above \$40 million exiting Q2 FY27 (up from \$15 million exiting Q1 as restated on this call); target above \$100 million paid CRR by fiscal year-end Source	The latest dollar figure is platform-wide, not exclusively AI; a growth rate without an AI dollar base is excluded.
Atlassian (TEAM)	No qualifying record established	No qualifying primary record established here.	The existing record contains adoption, qualitative commentary or no amount. This is not proof that all disclosures lack a qualifying figure.
JFrog (FROG)	No qualifying record established	No qualifying primary record established here.	The existing record contains adoption, qualitative commentary or no amount. This is not proof that all disclosures lack a qualifying figure.
UiPath (PATH)	Unverified	2026-03-11: AI product ARR (agentic, IDP and Maestro) nearly \$200 million Source	No complete primary-source four-quarter review is recorded under this rule.
Datadog (DDOG)	This record excluded	2026-08-06: revenue growth for non-AI customers in the high-20s percent against 36% total growth (share-of-growth framing); cohort share of revenue not stated Source	Share of growth is excluded. An earlier absolute amount or revenue share may qualify after primary-source review.
Dynatrace (DT)	No qualifying record established	No qualifying primary record established here.	The existing record contains adoption, qualitative commentary or no amount. This is not proof that all disclosures lack a qualifying figure.
PagerDuty (PD)	No qualifying record established	No qualifying primary record established here.	The existing record contains adoption, qualitative commentary or no amount. This is not proof that all disclosures lack a qualifying figure.
Okta (OKTA)	No qualifying record established	No qualifying primary record established here.	The existing record contains adoption, qualitative commentary or no amount. This is not proof that all disclosures lack a qualifying figure.

COMPANY	RECORD STATUS	DATED EVIDENCE	REASON
SailPoint (SAIL)	Qualifying record verified	2026-09-09: AI-driven ARR above \$70 million Source	Product grouping defined by SailPoint; not comparable to recognized revenue.
Zscaler (ZS)	Qualifying record verified	2026-05-26: AI Protect trailing-12-month bookings above \$100 million Source	Bookings include full contract value across the term; not annual revenue.
Netskope (NTSK)	No qualifying record established	No qualifying primary record established here.	The existing record contains adoption, qualitative commentary or no amount. This is not proof that all disclosures lack a qualifying figure.
CrowdStrike (CRWD)	No qualifying record established	No qualifying primary record established here.	The existing record contains adoption, qualitative commentary or no amount. This is not proof that all disclosures lack a qualifying figure.
SentinelOne (S)	No qualifying record established	No qualifying primary record established here.	The existing record contains adoption, qualitative commentary or no amount. This is not proof that all disclosures lack a qualifying figure.
Tenable (TENB)	No qualifying record established	No qualifying primary record established here.	The existing record contains adoption, qualitative commentary or no amount. This is not proof that all disclosures lack a qualifying figure.
Qualys (QLYS)	No qualifying record established	No qualifying primary record established here.	The existing record contains adoption, qualitative commentary or no amount. This is not proof that all disclosures lack a qualifying figure.
Palo Alto Networks (PANW)	Qualifying record verified	2026-09-01: Prisma AIRS ARR above \$100 million Source	Based on software NGFW credits allocated to AIRS on the booked quote; customers can consume credits across products.
Rubrik (RBRK)	No qualifying record established	No qualifying primary record established here.	The existing record contains adoption, qualitative commentary or no amount. This is not proof that all disclosures lack a qualifying figure.

COMPANY	RECORD STATUS	DATED EVIDENCE	REASON
Commvault (CVLT)	No qualifying record established	No qualifying primary record established here.	The existing record contains adoption, qualitative commentary or no amount. This is not proof that all disclosures lack a qualifying figure.
AvePoint (AVPT)	No qualifying record established	No qualifying primary record established here.	The existing record contains adoption, qualitative commentary or no amount. This is not proof that all disclosures lack a qualifying figure.

Three records kept apart: the prediction as registered, the current forecast, and the growth required to reach the prediction from the current anchor; a fourth line shows the forecast if it were restarted from the latest OpenRouter-based estimate

RECORD	2026 STARTING POINT	2032 HOURS, AW / ME	CROSSES 7.0 TRN, AW / ME	GROWTH A YEAR: TO CROSS · TO AW / ME	FOUR- COMPANY REVENUE: 2029 AW / ME · 2032 AW / ME
The prediction Registered October 8, 2026; fixed, moves only by a new registration in D2.	35M a day	19.0 / 340 trn	2032 / 2030	2.86x · 3.38x / 5.46x	\$40 / \$49bn · \$71 / \$124bn
The current forecast The anchor has not been re-based since launch, so the current forecast equals the prediction.	35M a day	19.0 / 340 trn	2032 / 2030	2.86x · 3.38x / 5.46x	\$40 / \$49bn · \$71 / \$124bn

RECORD

2026
STARTING
POINT

2032 HOURS,
AW / ME

CROSSES
7.0 TRN,
AW / ME

GROWTH
A YEAR:
TO CROSS
· TO AW /
ME

FOUR-
COMPANY
REVENUE:
2029 AW /
ME · 2032
AW / ME

If restarted from the latest OpenRouter-based estimate	55M a	30.1 / 537	2031 /	2.65x ·	\$41 /
What the current forecast would read if it were restarted	day	trn	2030	3.13x /	\$50bn ·
from the latest OpenRouter-based estimate, growth rates		demand ·		5.06x	\$79 /
held; not a re-base, the rule has not fired. The whole		30.1 / 357			\$133bn
hours path scaled by the ratio of the latest pulse to the		trn delivered			
anchor, each year’s growth preserved, then held at the		under the			
published capacity path where demand exceeds it, as		capacity rule			
every case on the model page is; destinations are the					
scaled demand, and crossing years and revenue read from					
the delivered hours. The growth columns are the growth					
from this start to the registered destinations, a different					
calculation from the scaled path’s own rates.					

The prediction is fixed; the current forecast moves when the anchor is re-based under the A4 rule, with the published growth rates held; the required growth is recomputed from the current anchor at every edition. The crossing is read against 7.0 trillion paid hours a year (A5). AW is the Agentic Workforce, ME the Machine Economy; growth multipliers are a year, 2026 to 2032.

Growth a year to the crossing by starting point, week to October 18, 2026: from the provider estimate (17 million a day) 3.21x, from the original starting point (35 million a day) 2.86x, from the OpenRouter-based estimate (55 million a day) 2.65x. The forecast is graded from the anchor; the other two show what the choice of starting point does to the requirement.

The crossover and the backend thesis can receive different verdicts: four outcomes and how each would be read

OUTCOME	WHAT IT WOULD MEAN FOR THE TWO CLAIMS	HOW WE WOULD READ IT
Machine hours cross paid human hours by 2032, and durable paid backend demand develops	Both predictions hold: more machine hours and lasting paid infrastructure use.	The remaining question is which providers earn the revenue.
Machine hours cross; durable paid backend demand does not develop	The hours prediction holds, but revenue assumptions fail: free tiers, efficiency, lower prices and bundled services limit paid demand.	More activity, limited provider revenue: the hours prediction succeeds while the revenue prediction fails. Each requires separate evidence.
Machine hours do not cross; durable paid backend demand develops anyway	The timing prediction fails; the infrastructure thesis holds through the core business and retained demand.	The near-term claim can still grade as a hit on reported growth while the hours prediction misses.
Machine hours do not cross, and durable paid backend demand does not develop	Both propositions fail.	The thesis is wrong, and the record says so at each deadline.

The matrix is the forecast record’s judging rule, two claims by two outcomes: each cell names the evidence that would put us in it, so no single event, IPO or acquisition reads as confirmation. Activity that grows mainly through retries and low-value experimentation is a quality diagnostic on the first claim, read with completed

tasks and resources per accepted result, not a fifth cell. The original prediction, its measurement rules and its update policy are stated beside it in D2.

D3 • Disclosures

The full text of the site's disclosures: research commentary, not investment advice; [Skycatcher's](#) investment in Clockwork Labs, the company behind Spacetime, which also serves this site; positions

and conflicts, market comparisons, data, third-party material and privacy.

Research commentary. Machine Hour is research commentary published by [Skycatcher](#), LLC, an exempt reporting adviser. We publish it on a regular schedule to the general public to share an idea. It does not solicit investors for any [Skycatcher](#) fund. It is published for general research and discussion and does not consider any reader's circumstances. It is not an offer or solicitation of any security or fund, not a recommendation and not investment advice, and it is not intended for distribution where that would be unlawful.

Forward-looking statements. The scenarios, the machine hours model, the revenue paths for Cloudflare, MongoDB, Snowflake and Datadog, the bold bets and every figure marked modeled are [Skycatcher](#) estimates built on stated assumptions. They are not company guidance, consensus or assurances of future results, and the companies named have not reviewed them. They can be wrong. A bold bet is a dated prediction with a public measure; its grade records whether the prediction landed and is not a record of trading, returns or advice, and no bet is an instruction to buy or sell anything. We grade the bets at their deadlines and publish every grade, hit or miss; we are under no duty to update any other statement. We fix and date the scenario inputs at each edition and grade the tracker's hours gauge each quarter on the machine hours model; the crossover years move with that grading. Several figures and the model page carry sliders: a case a reader draws with them is that reader's own input, not a [Skycatcher](#) estimate, and it is not stored or published.

Framework and company cases. State, Work, Applications, Intelligence and Computation are research classifications of what a company sells. The roster, the four bellwethers and every named company case are illustrations of the argument, not holdings disclosures, a portfolio, a model portfolio or a screened investment basket.

Positions and conflicts. [Skycatcher](#), its affiliates, principals and clients may hold long or short positions in any public company named on this site, including the 32 companies of the backend universe and the four bellwethers, and may buy or sell them before, during and after publication without notice. No company named on this site pays [Skycatcher](#) in any form for its coverage. [Skycatcher](#), its affiliates and principals hold venture positions in some of the private companies mentioned, including Clockwork Labs, the company behind Spacetime. This site is served by a Spacetime module on that company's Maincloud service, and subscriber email addresses are stored there and processed by Beehiiv to send the monthly email. Our operating experience with Spacetime and the vendor's published benchmarks are not independent research, and the competitors named in those benchmarks have not verified them.

Market comparisons. Price comparisons use split- and dividend-adjusted closes in equal-weight baskets. We defined the baskets in 2026 and applied them back to January 2022, so they carry hindsight selection. They are hypothetical: not an investable index, not an actual account, and no fees, taxes or trading costs are deducted. Historical observations are not a guarantee of future returns. Consensus figures are the analyst mean carried by Financial Modeling Prep on the date stated and can change as analyst coverage changes.

Data and dates. Company financial observations are Financial Modeling Prep's standardized quarterly statements, each tagged with its reporting period and filing date and cut at the filing date, refreshed each trading night; coverage is shown, and the SEC filings remain the primary source a reader can check. Market prices,

balance-sheet components and trailing revenue can have different dates; valuation approximations disclose their basis. Missing observations remain missing. Non-US companies are measured in their listing currency unless a conversion is stated.

Third-party material. Names and marks identify their owners and do not imply endorsement. Nasdaq-100 is a registered trademark of Nasdaq, Inc., used here only to identify the benchmark; Nasdaq does not sponsor, endorse or review this site. Company figures are summarized in original illustrations with links to the underlying publications. Names and source links do not imply permission, endorsement or independent verification. Rights questions can be sent through the About page. Links to other sites are provided for reference; we do not control their content and are not responsible for it. Analyst consensus, market values and the balance-sheet inputs to valuation approximations come from Financial Modeling Prep (FMP); company financials from SEC filings; token volumes from OpenRouter; software-production counts from the GitHub Search API; question counts from Stack Overflow (CC BY-SA 4.0); task-horizon observations from METR.

Privacy. The subscribe form records your email address, signup time and the form used. [Skycatcher](#), LLC handles subscription requests. Addresses are stored by Spacetime and shared with Beehiiv for email delivery. New subscriptions require email confirmation before they become active in Beehiiv. We do not automatically reactivate an unsubscribed address. Use the unsubscribe link in an email to stop updates, or the About contact link to request access or deletion. We may retain a suppression record to prevent further mail. Vercel and Spacetime serve these pages and may process request information, including IP addresses, in their service logs. The chart library is served with the site. This site code sets no cookies or analytics tags; it stores your theme preference in your browser. Beehiiv email tracking and processing depend on the publication settings and its privacy notice. Processing locations and retention periods depend on the providers and applicable obligations; no fixed location or deletion period is promised here.

Accuracy and liability. We take care with the data and the arithmetic and link each figure to its source, but we do not warrant that anything here is accurate, complete or current; third-party data is shown as received. To the extent the law allows, [Skycatcher](#), LLC accepts no liability for loss arising from use of or reliance on this site. Corrections reach us through the About page, and we publish each one, dated, in the monthly update.

Jurisdiction and contact. This site is published from the United States. Questions, corrections, rights-holder notices, grading disputes and privacy requests go through the contact link on the About page. Disclosures last reviewed October 8, 2026.

Appendix E — Company presentations: findings summarized in our own illustrations

E1 • Cloudflare, June 9, 2026

The June 9, 2026 slides show how Cloudflare connects its network and developer platform to agent activity. The most useful evidence concerns growth in requests and how contracts turn those requests into revenue.

Its agents-per-worker example is a planning assumption. It is context for our scenarios, not a measured floor under global adoption. Our original illustrations retain the stated metrics and link to the company source.

Where and when work runs: one room and batch, then desktop, then everywhere, then anywhere and continuous

- 01

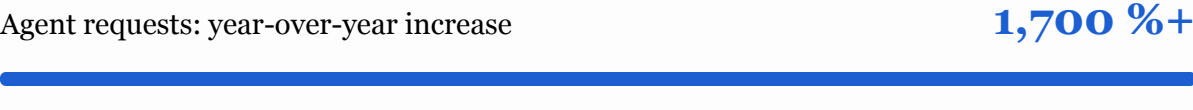
Work becomes less tied to one location
- 02

Agents can keep tasks running between human requests

Skycatcher illustration based on the source below. Conceptual summary, not a measured quantitative series.

Source: [Cloudflare, Investor Day deck, June 9, 2026](#)

AI agent requests a day on Cloudflare’s network, June 1, 2025 to May 31, 2026



Skycatcher illustration based on the source below. Values are company-reported, with their stated scope; intermediate points are not invented.

Source: [Cloudflare, Investor Day deck, June 9, 2026](#)

Cloudflare’s planning arithmetic: a billion knowledge workers, ten agents each, and the CPUs that would take

Knowledge workers assumed	1 billion
Agents per worker assumed	10
Agents per CPU assumed	10

Skycatcher illustration based on the source below. Values are company-reported, with their stated scope; intermediate points are not invented.

Source: Cloudflare, Investor Day deck, June 9, 2026

Four gauges of coding output, year on year, as Cloudflare read them



Skycatcher illustration based on the source below. Values are company-reported, with their stated scope; intermediate points are not invented.

Source: Cloudflare, Investor Day deck, June 9, 2026

How Cloudflare bills: the cap model, with requests as the unit of measure

01 Customer contract sets a monthly cap

02 Request volume influences the bill

Skycatcher illustration based on the source below. Conceptual summary, not a measured quantitative series.

Source: Cloudflare, Investor Day deck, June 9, 2026

E2 · Datadog, February 12, 2026

The February 12, 2026 slides describe a move from monitoring toward systems that can respond to problems. They show rapid growth in AI-related records and the contribution of AI-focused customers.

The pricing example combines more volume with a different storage mix. It should not be read as a like-for-like price reduction. The distinction matters when using it to stress the revenue model.

Datadog's three eras: monitoring, observability, autonomy

01 Monitor a system

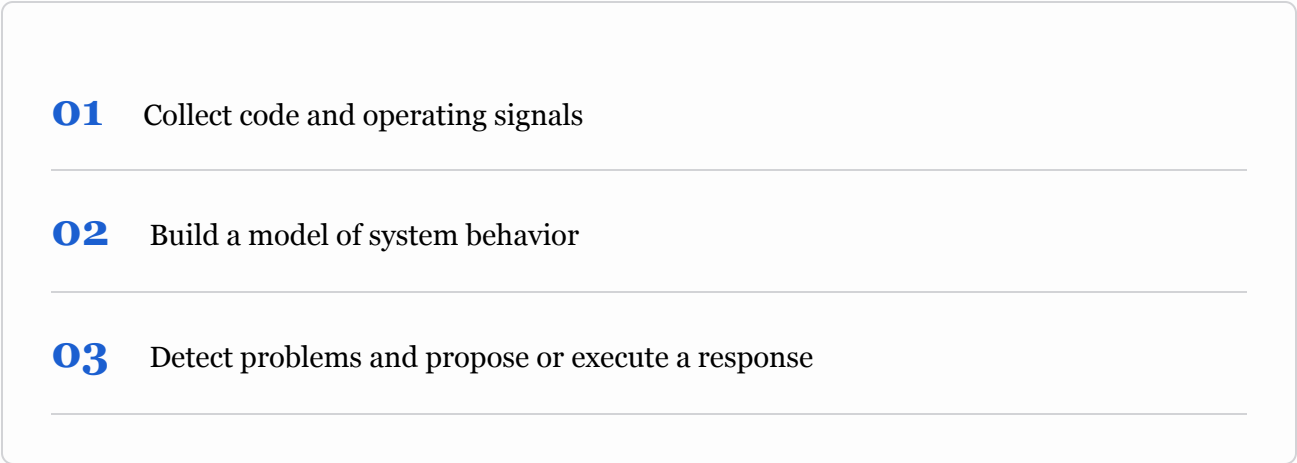
02 Understand its behavior

03 Respond within authorized limits

Skycatcher illustration based on the source below. Conceptual summary, not a measured quantitative series.

Source: Datadog, Investor Day deck, February 12, 2026

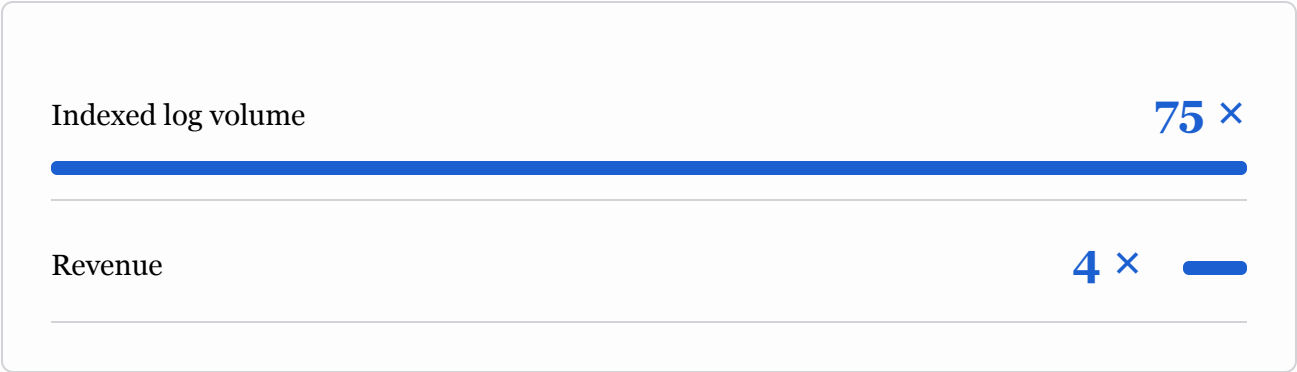
Datadog’s path to autonomous operations: telemetry into a world model of the system, out to remediation without a person



Skycatcher illustration based on the source below. Conceptual summary, not a measured quantitative series.

Source: [Datadog, Investor Day deck, February 12, 2026](#)

One customer, 2021 to 2025: log volume about 75x, unit price down, revenue about 4x



Skycatcher illustration based on the source below. Values are company-reported, with their stated scope; intermediate points are not invented.

Source: [Datadog, Investor Day deck, February 12, 2026](#)

AI Observability spans: July 2025–January 2026



Skycatcher illustration based on the source below. Values are company-reported, with their stated scope; intermediate points are not invented.

Source: [Datadog, Investor Day deck, February 12, 2026](#)

Datadog MCP server tool calls, July 2025 to January 2026

01 Agents call Datadog through MCP

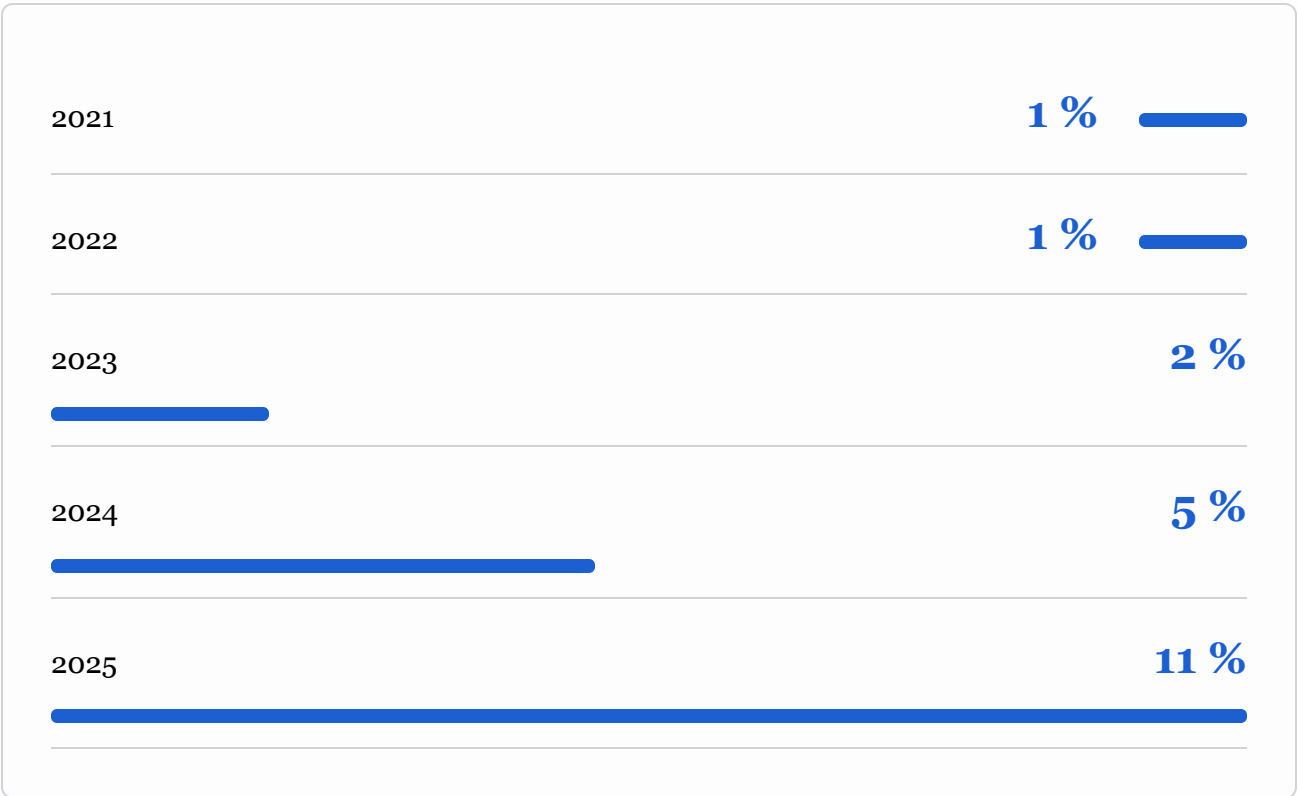
02 The company reports rapid growth

03 No numeric series is transcribed here

Skycatcher illustration based on the source below. Conceptual summary, not a measured quantitative series.

Source: [Datadog, Investor Day deck, February 12, 2026](#)

AI-native customers’ share of Datadog revenue, 2021 to 2025



Skycatcher illustration based on the source below. Values are company-reported, with their stated scope; intermediate points are not invented.

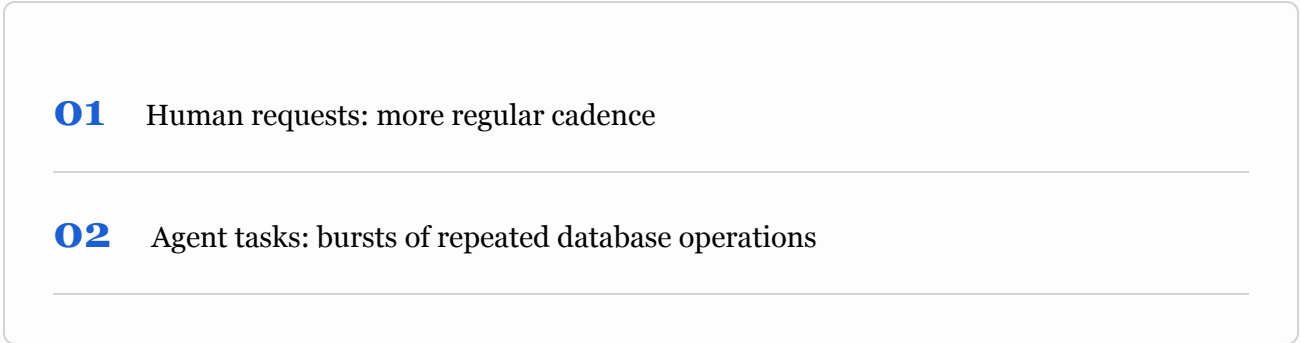
Source: [Datadog, Investor Day deck, February 12, 2026](#)

E3 • MongoDB, September 29, 2026

The September 29, 2026 slides show the scale of Atlas, the connection of agent tools to its database and growth among customers building AI products.

Approximately 40% of Atlas recurring revenue came from customers with at least one AI use case. That is a customer classification, not an AI-revenue share. The fast-growing AI-focused customer group was still a small part of Atlas revenue.

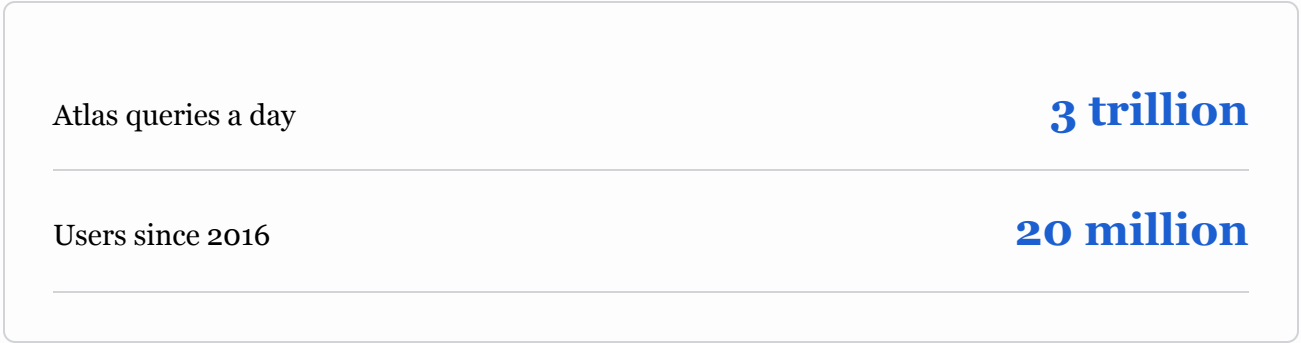
User-driven workloads against agentic workloads, as MongoDB drew them



Skycatcher illustration based on the source below. Conceptual summary, not a measured quantitative series.

Source: [MongoDB, Investor Day deck, September 29, 2026, filed as Exhibit 99.1 to its Form 8-K](#)

Atlas: about 20 million users since 2016 and 3 trillion queries a day



Skycatcher illustration based on the source below. Values are company-reported, with their stated scope; intermediate points are not invented.

Source: [MongoDB, Investor Day deck, September 29, 2026, filed as Exhibit 99.1 to its Form 8-K](#)

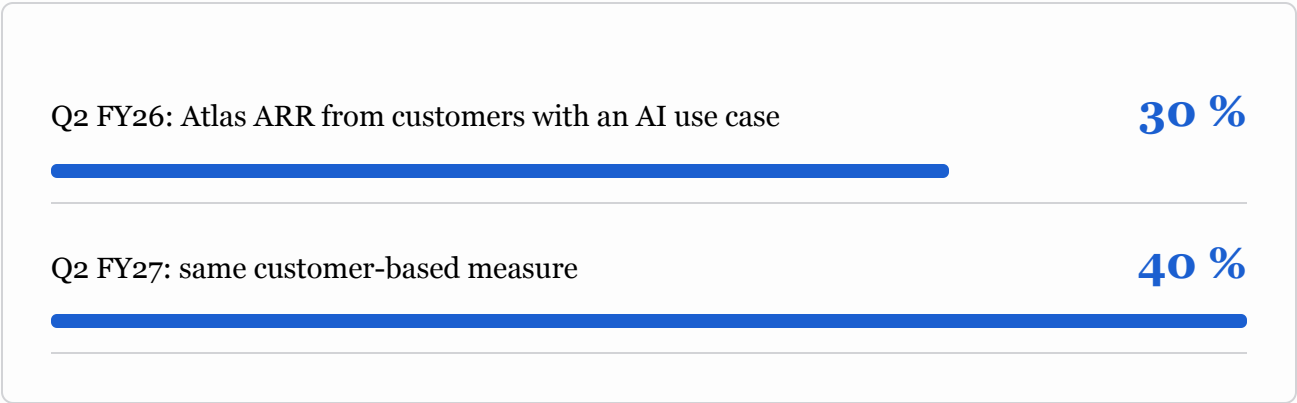
The managed MCP server, seven weeks from launch, and the agents that call it



Skycatcher illustration based on the source below. Values are company-reported, with their stated scope; intermediate points are not invented.

Source: [MongoDB, Investor Day deck, September 29, 2026, filed as Exhibit 99.1 to its Form 8-K](#)

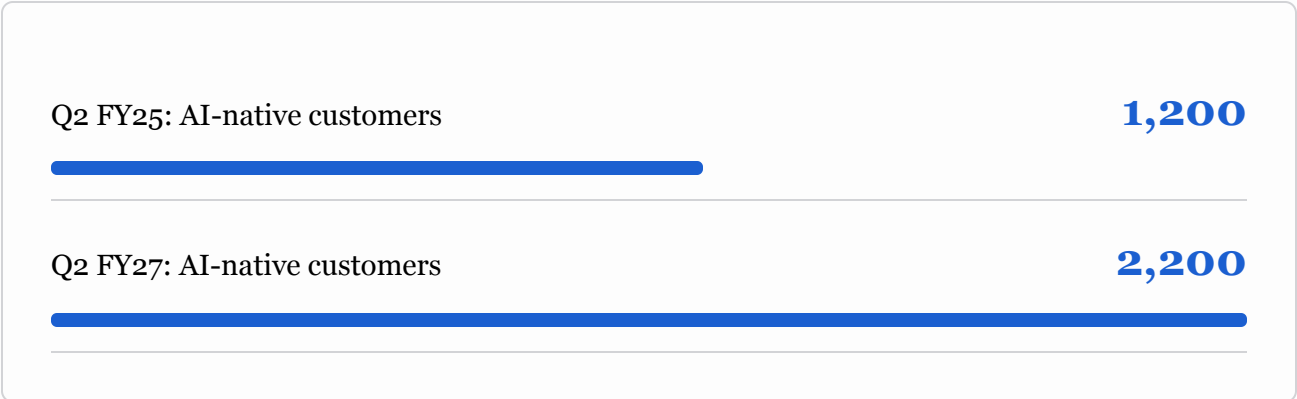
Share of Atlas ARR from customers with at least one AI use case



Skycatcher illustration based on the source below. Values are company-reported, with their stated scope; intermediate points are not invented.

Source: MongoDB, Investor Day deck, September 29, 2026, filed as Exhibit 99.1 to its Form 8-K

AI-native customer counts, Q2 FY25 to Q2 FY27



Skycatcher illustration based on the source below. Values are company-reported, with their stated scope; intermediate points are not invented.

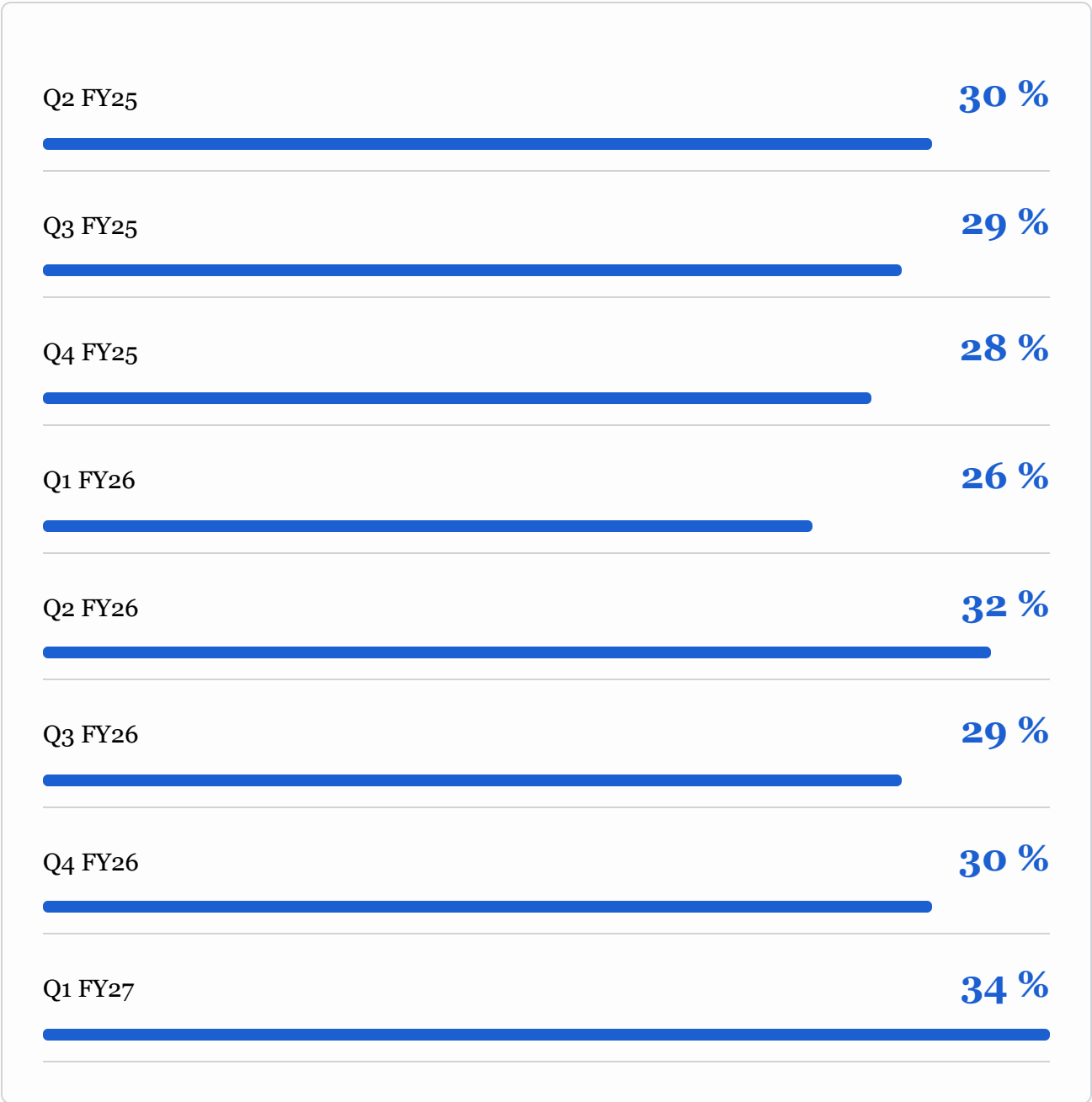
Source: MongoDB, Investor Day deck, September 29, 2026, filed as Exhibit 99.1 to its Form 8-K

E4 • Snowflake, June 2, 2026

The June 2, 2026 slides show product growth, existing-customer expansion and examples of agents working with governed business data.

The task example compares answers with and without additional runtime context. Its accuracy and cost figures belong to that setup. They demonstrate a possible benefit of supplying relevant context, not a universal advantage over every alternative.

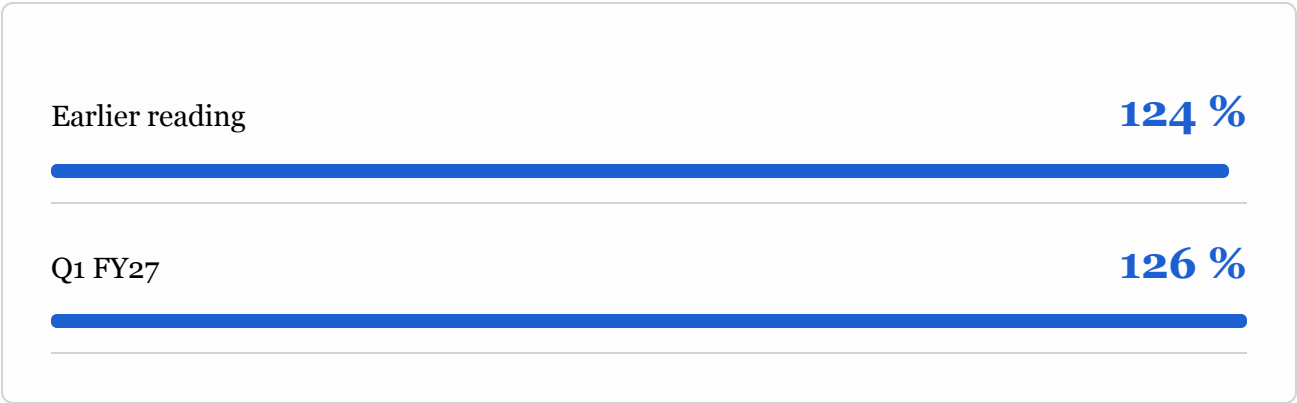
Product revenue growth by quarter, Q2 FY25 to Q1 FY27



Skycatcher illustration based on the source below. Values are company-reported, with their stated scope; intermediate points are not invented.

Source: [Snowflake, Investor Day deck, June 2, 2026](#)

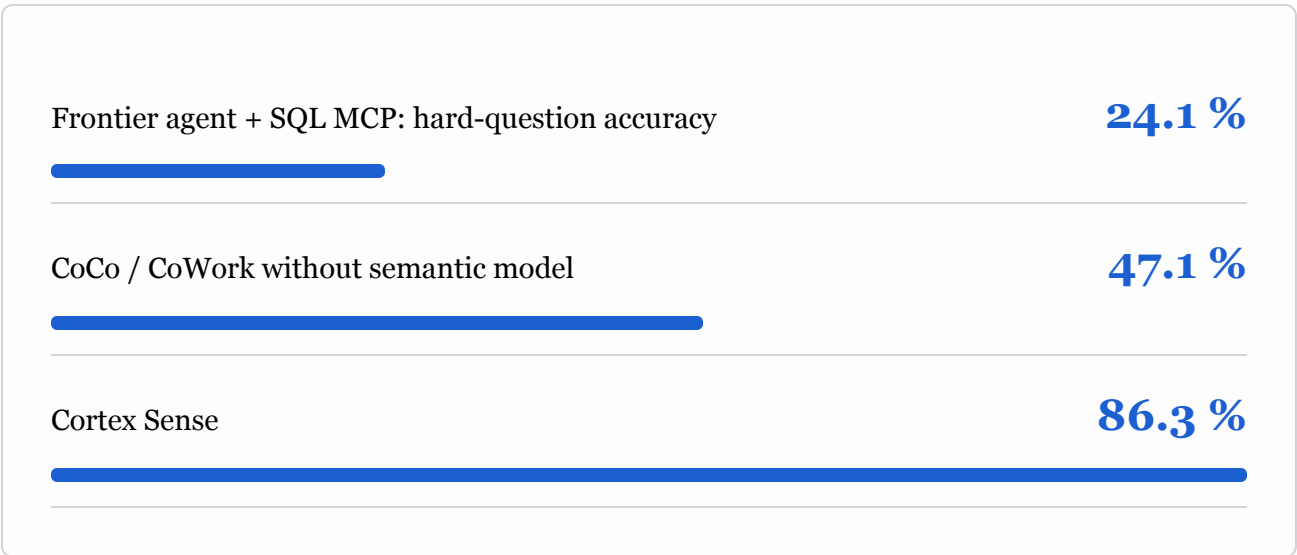
Net revenue retention, Q1 FY26 to Q1 FY27



Skycatcher illustration based on the source below. Values are company-reported, with their stated scope; intermediate points are not invented.

Source: [Snowflake, Investor Day deck, June 2, 2026](#)

What an agent prompt costs and how often it is right, on Snowflake’s sample



Skycatcher illustration based on the source below. Values are company-reported, with their stated scope; intermediate points are not invented.

Reported average cost per prompt for the same three configurations: \$1.76, \$0.49 and \$0.59. Vendor evaluation; workload and configuration affect comparability.

Source: [Snowflake, Investor Day deck, June 2, 2026](#)

About

Machine Hour is public research from [Skycatcher](#), an investment firm founded in 2015. We study the Internet by using the products being built on it. Putting agents to work ourselves led us to the question behind this site: how much work can machines perform, and what infrastructure will that work need?

What we study

We study AI agents doing work for people and the infrastructure that supports them. Our agents read filings, query data, write software and check results. Using them helps us see their strengths and failures; it does not replace broader evidence.

A machine hour is one agent actively working for one hour. It measures activity, not the value or quality of the result. We estimate that activity, explore how it could grow and track whether it creates lasting demand for software infrastructure.

How to follow the research

Start with the [summary](#). Five parts develop the thesis. Use the [model](#) to test the assumptions and the [tracker](#) to follow the evidence. The [appendices](#) explain sources, methods and limitations.

The thesis and human-hours prediction were published on October 8, 2026. Predictions keep their original terms. The monthly update explains what the evidence supports, what challenges our view and what remains unknown.

Write to us

For questions, corrections, better data or privacy requests, [write to Skycatcher](#) (machine@skycatcher.xyz).

Disclosures

[Skycatcher](#), LLC is an exempt reporting adviser. This site is published for general research and discussion. It is not an offer of any fund, a solicitation, a recommendation or investment advice. It does not consider individual circumstances and is not intended for distribution where that would be unlawful.

Scenario outputs are our estimates on stated assumptions, not company guidance. Reported data retain their sources and dates. The index comparisons are hypothetical baskets defined in 2026 and applied back to 2022.

[Skycatcher](#), its affiliates, principals and clients may hold positions in any public company named, including the 32 backend companies and the four company cases, and may trade them without notice. [Skycatcher](#) is an investor in Clockwork Labs, the company behind Spacetime.

This site runs on a Spacetime module and stores subscriber addresses there. Beehiiv delivers the monthly email. See the [full disclosures and privacy note](#).

Tracker • Machine Hour

Where things stand

ACTIVITY

55M machine hours a day • Skycatcher estimate using OpenRouter data

Alternative provider-by-provider estimate: 17 million a day. Neither is a direct count. The original 2026 forecast starts at 35 million a day; methods are in Tokens.

CHANGE	AS OF	NEXT
OpenRouter estimate +23.2% in four weeks; provider estimate +6.1% on the month	week to October 18, 2026	CHECKPOINT the path needs 67 to 85M a day by June 30, 2027

MONETIZATION

Backend revenue growth 20.7%, re-accelerating

Median growth of the backend basket, against Computation at 70.2% and the Nasdaq 100 at 14.2%. Latest activity reading: MongoDB, AI-native customers on Atlas: 2,200+ (Q2 FY27). The company panels test whether activity becomes paid use.

CHANGE	AS OF	NEXT
+1.5 points on the quarter before (19.2%)	quarter to June 30, 2026	CHECKPOINT the quarter to September 30, 2026, reported in the following weeks

VALUATION

Four-company calculations imply 21 to 53% annual growth; our scenarios use 24 to 42%

Above both scenario growth paths: Cloudflare. The company chart compares each valuation with its 2026–2031 scenarios.

CHANGE	AS OF	NEXT
trailing multiple +14% in a month, from 7.5× on September 22, 2026	close of October 22, 2026	CHECKPOINT the implied growth is recomputed daily; the next test is the quarter's consensus revisions

THESIS TESTS

11 on the record, 11 open

Graded once each at its deadline, the grade published beside the bet; the first grade falls on June 30, 2027. Capability: Claude Opus 4.6 (Inspect) finishes tasks of 719 expert minutes half the time, released February 5, 2026.

CHANGE	AS OF	NEXT
capability 12.0 expert hours at 50%, from 5.9 expert hours (December 11, 2025)	bets dated October 8, 2026	CHECKPOINT first grade on June 30, 2027

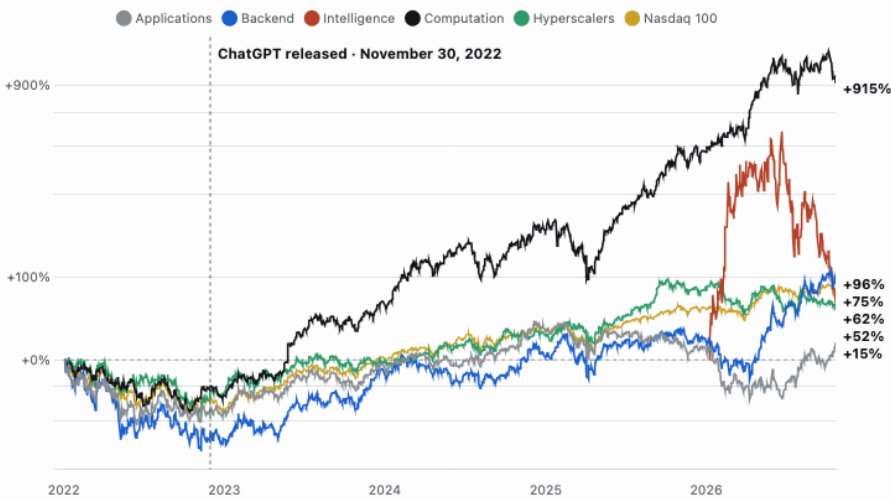
A1 The indices

DAILY Figures 0.3 · 3.1 · 4.1 in the essay →

Hypothetical equal-weight baskets defined in 2026 and applied back to 2022; hindsight-selected, no fees, taxes or trading costs, not an investable index.

EQUAL-WEIGHT INDICES · CHANGE SINCE JANUARY 3, 2022

The layers of the stack



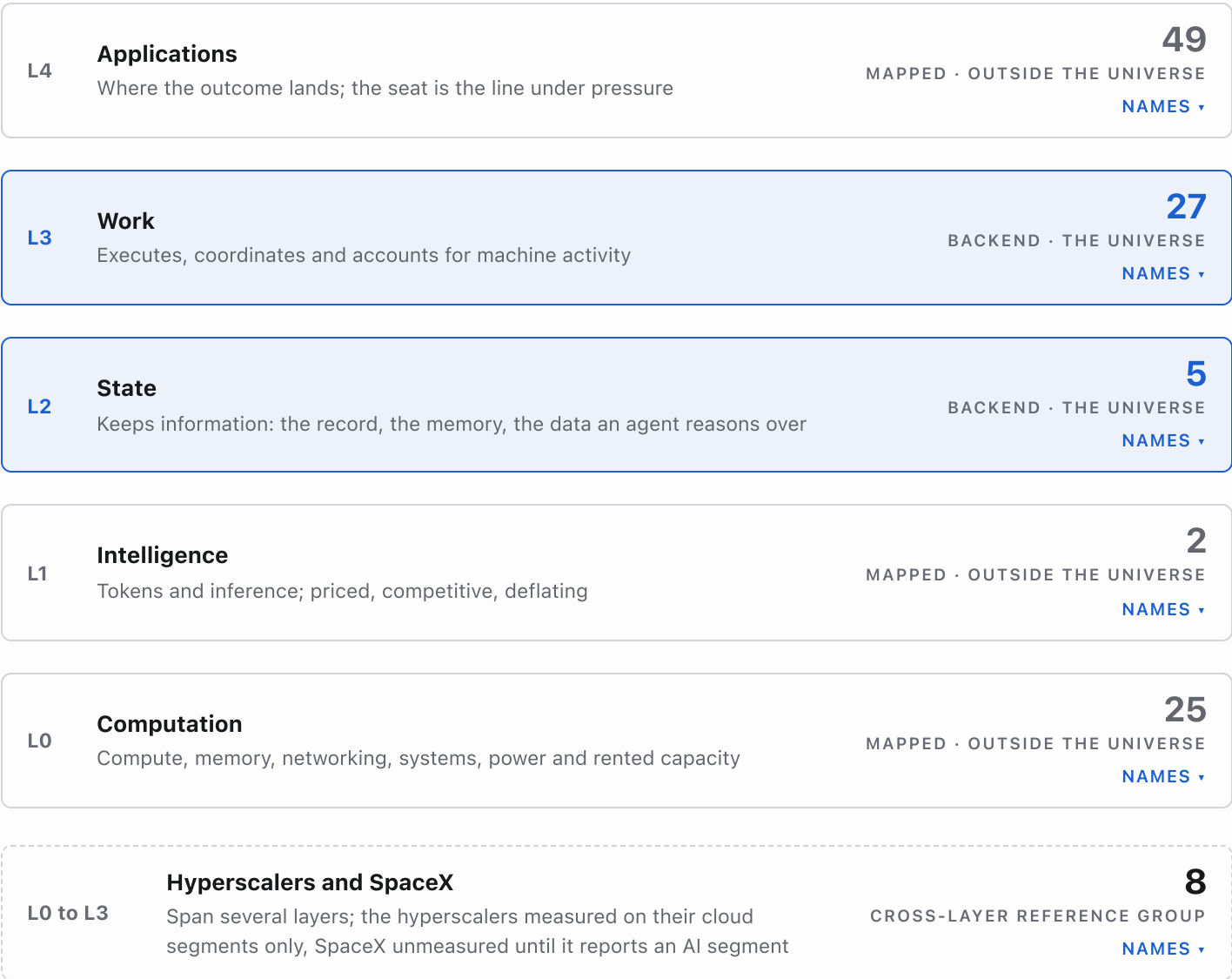
Latest: October 22, 2026 · Applications +15% · Backend +96% · Intelligence +62% (since January 8, 2026) · Computation +915% · Hyperscalers +52% · Nasdaq 100 +75%. Members: Applications 49, Backend 32, Intelligence 2, Computation 25, Hyperscalers 7.

One line per layer of [Figure 3.1](#), with State and Work drawn together as the backend, members as the [Appendix C1](#) roster stands today. Equal weight, reset daily, chained from January 3, 2022, each company's return in its own listing currency. The backend line is the published 32-name index; the other four layer lines follow the roster as it changes and are context. The Nasdaq 100 is the benchmark, rebased to the same day. The axis is the change since the chosen start on a log scale, so equal distances are equal percentage moves; the buttons above rebase every line, and clicking a legend entry hides that layer. This is [Figure 4.1](#) in Part Four. Intelligence has two members, both listed in 2026, so its line starts at 0% on its first trading day, and a Hong Kong holiday carries the prior level.

► [Method, baskets and entrants](#)

The five-layer stack of machine labor

PUBLIC PURE PLAYS PER LAYER · THE UNIVERSE IS L2 AND L3



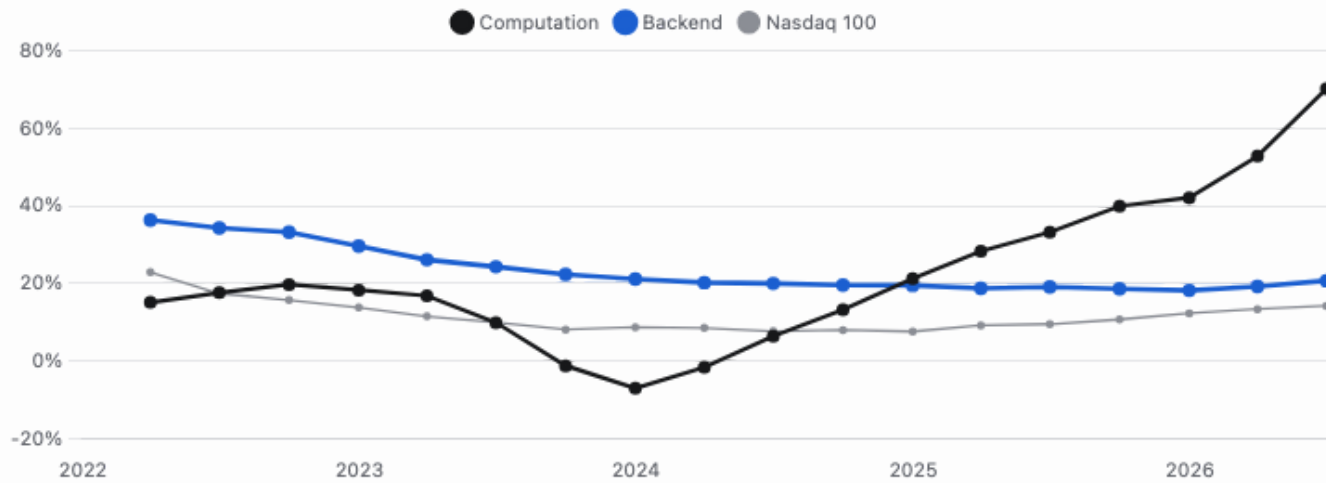
The October 1, 2026 map groups public companies by the services they provide. State and Work form the backend research universe. The other layers provide context; hyperscalers and SpaceX span several layers. Watchlist companies are listed but excluded from the count. The full roster is in Appendix C1. This is [Figure 3.1](#) in Part Three.

A2 Fundamentals

QUARTERLY Figures 3.4 · 3.6 · 3.7 · 4.2 in the essay →

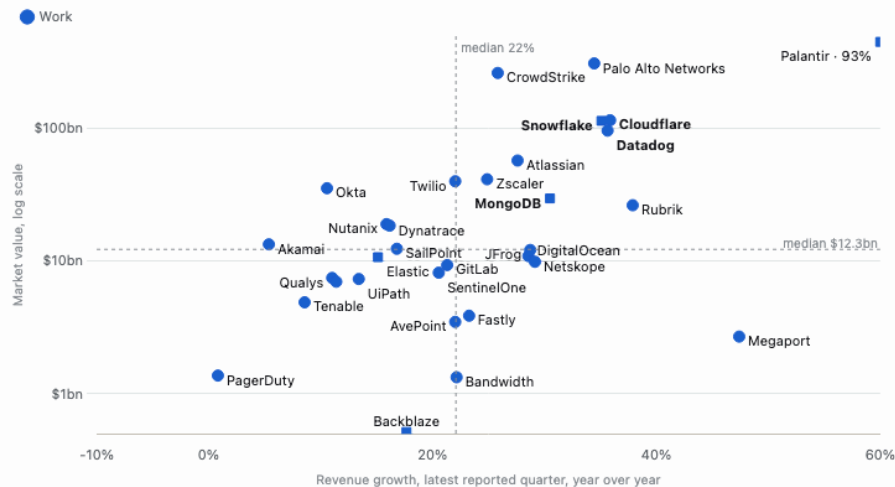
BACKEND RE-ACCELERATING Backend revenue growth 20.7% against Computation 70.2% in the quarter to June 30, 2026, from 19.2% and 52.8% the quarter before; the gap has widened against the backend for 10 quarters running. The Nasdaq 100 grows 14.2%.

Revenue growth



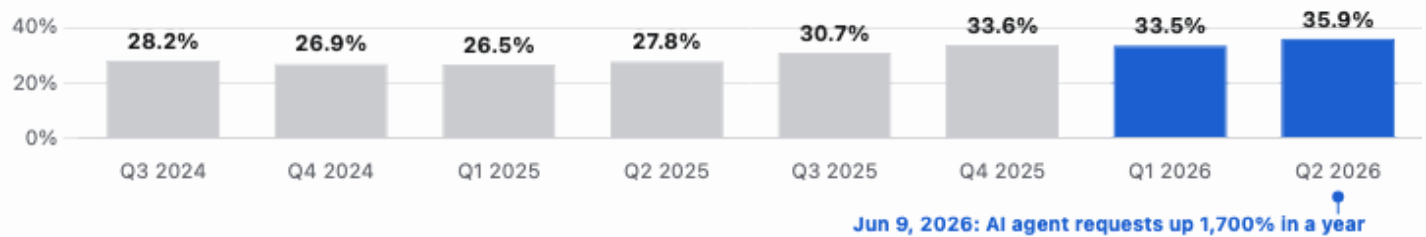
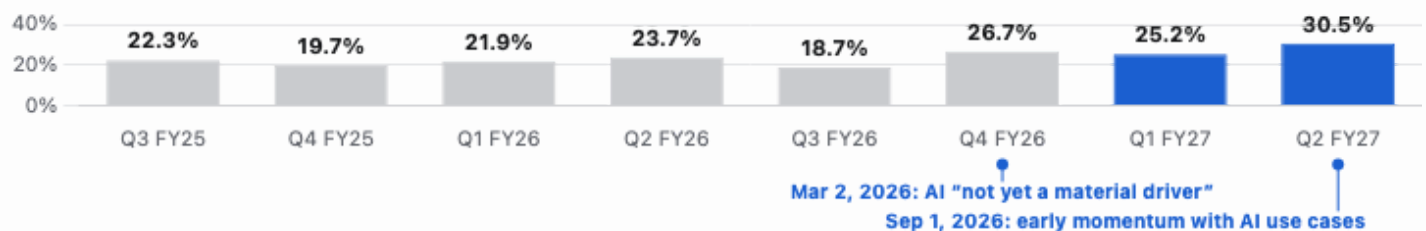
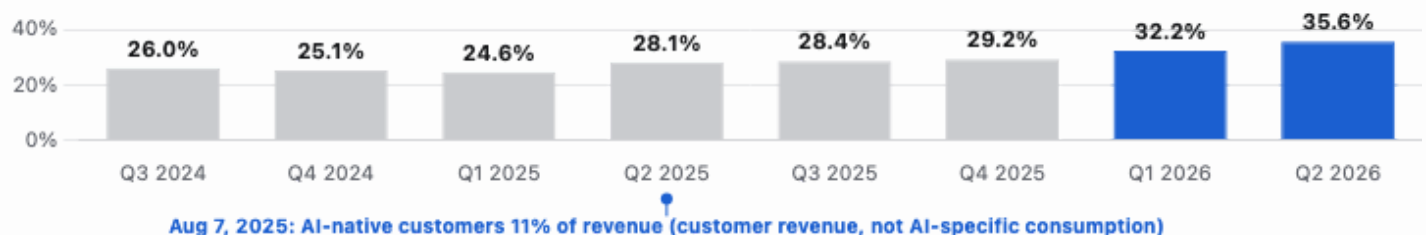
Median growth in trailing-twelve-month revenue, one point per calendar quarter. This is [Figure 4.2](#) in Part Four.

The public universe



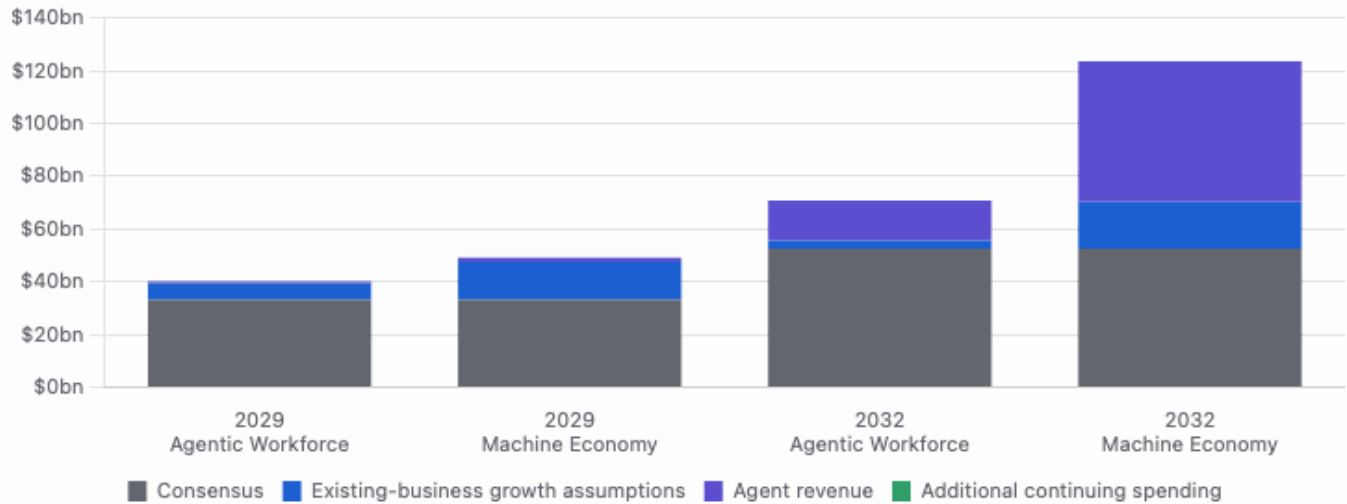
Latest-period revenue growth against market value at the October 22, 2026 close, log scale. State as squares and Work as circles. Dashed lines are the medians. Growth above 60% is pinned at the edge. Pick a layer above to see it alone. Hover a dot for the name. This is [Figure 3.4](#) in Part Three.

Four company cases

Cloudflare · reported through Q2 2026 · ended June 30, 2026**MongoDB** · reported through Q2 FY27 · ended July 31, 2026**Snowflake** · reported through Q2 FY27 · ended July 31, 2026**Datadog** · reported through Q2 2026 · ended June 30, 2026

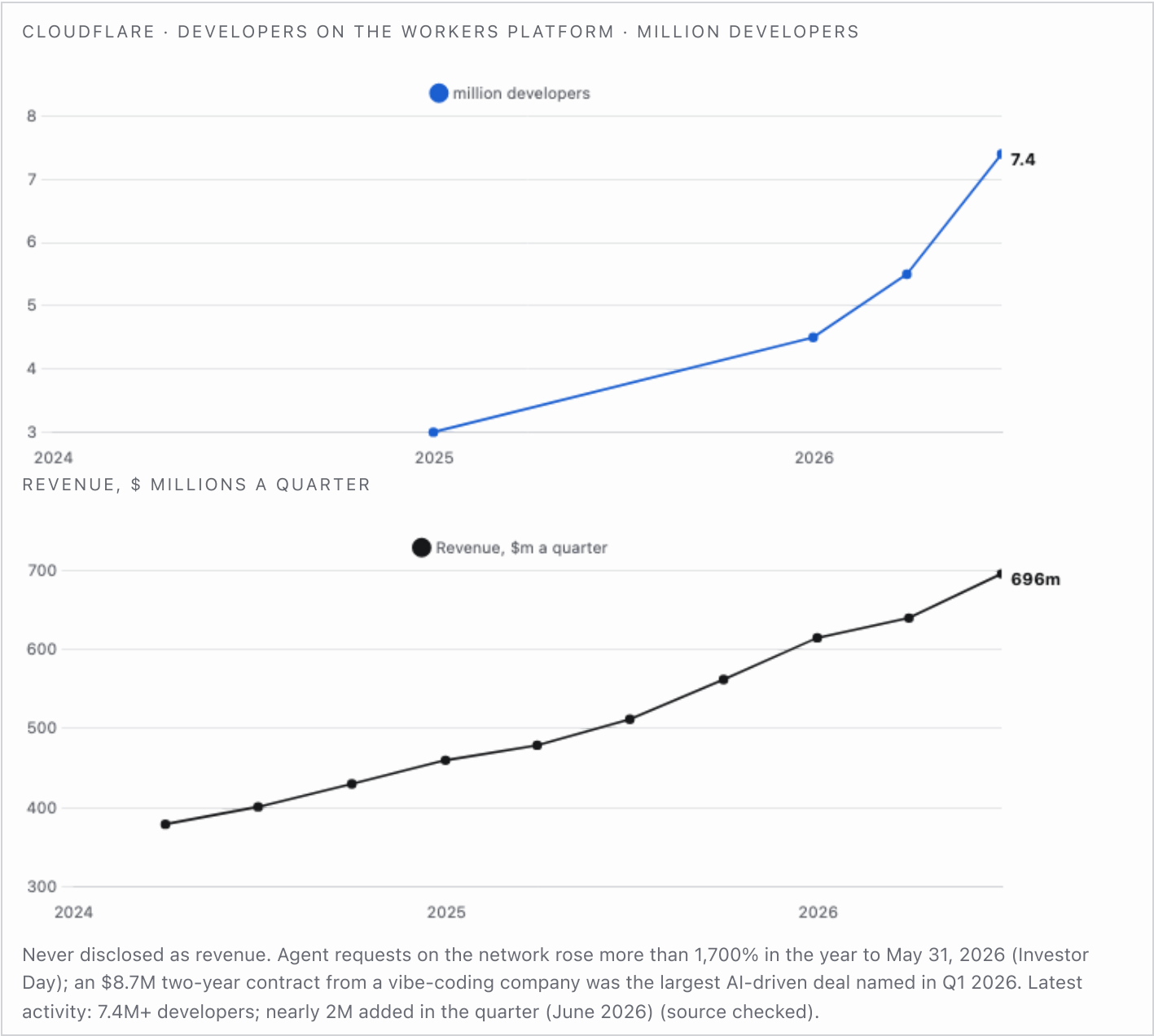
Revenue growth, year over year, from the filings, labeled by each company's fiscal quarter. The last two quarters are blue; the dated lines are management's AI statements on the calls. This is [Figure 3.6](#) in Part Three.

From consensus to the scenario

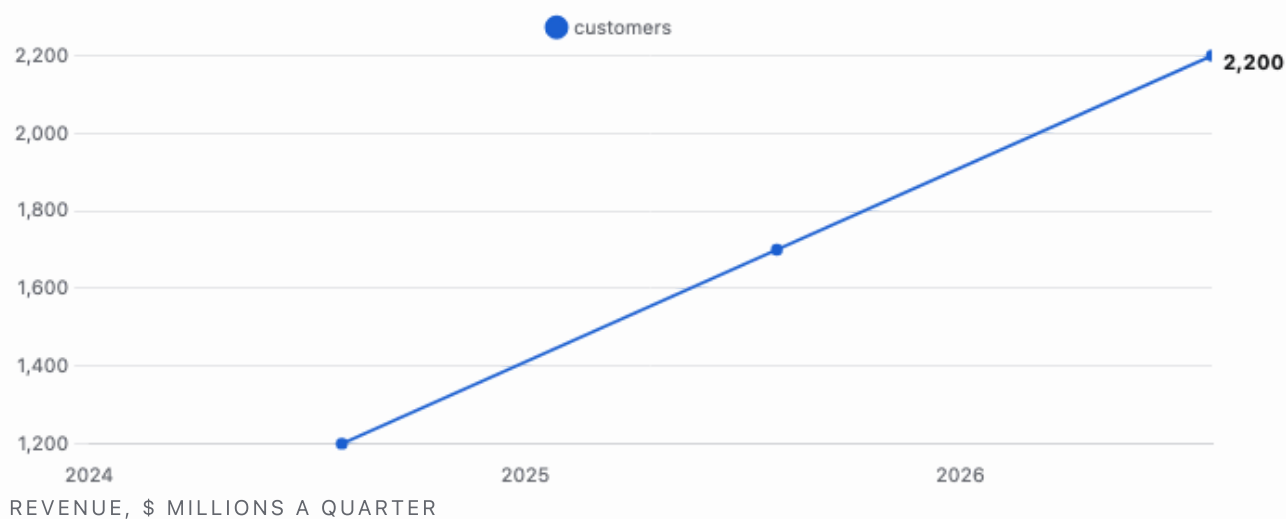
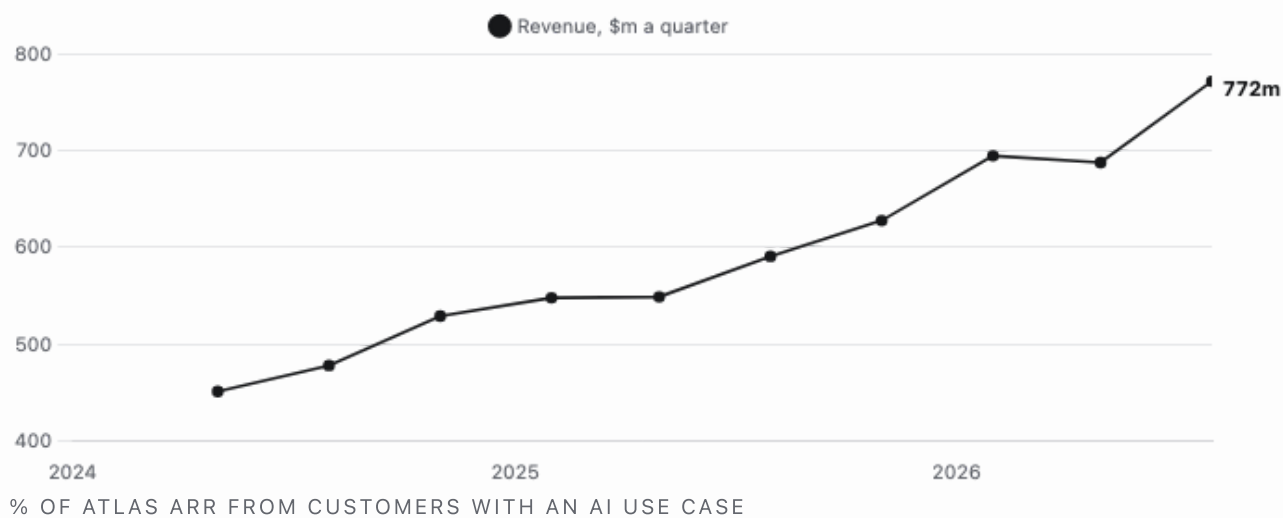
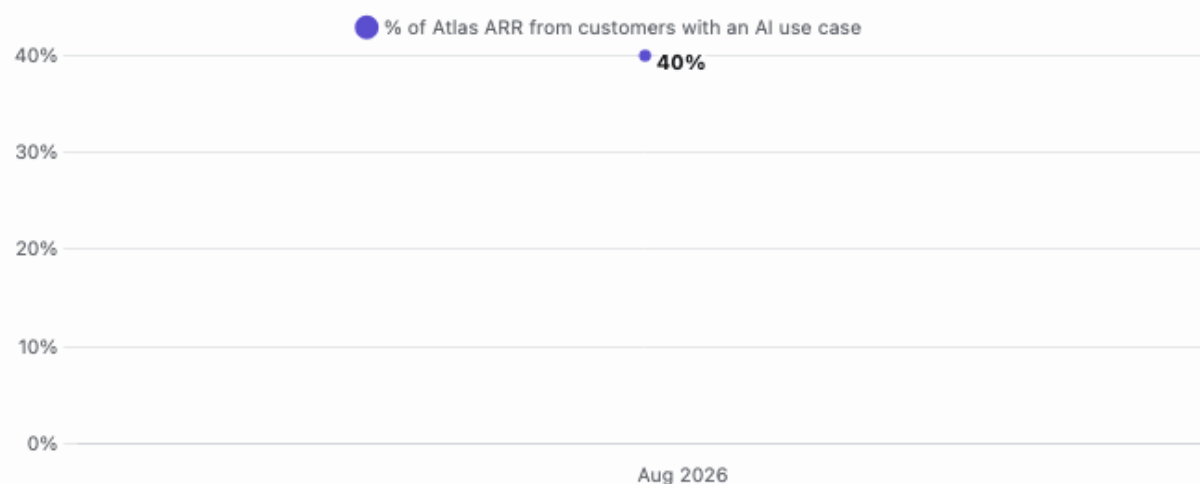


2029 · Agentic Workforce: consensus \$33.1bn, existing-business assumptions +\$6.3bn, agent revenue +\$0.9bn = \$40.2bn (88% of the gap comes from existing-business assumptions). 2029 · Machine Economy: consensus \$33.1bn, existing-business assumptions +\$14.4bn, agent revenue +\$1.7bn = \$49.1bn (90% of the gap comes from existing-business assumptions). 2032 · Agentic Workforce: consensus \$52.4bn, existing-business assumptions +\$3.3bn, agent revenue +\$15.0bn = \$70.7bn (18% of the gap comes from existing-business assumptions). 2032 · Machine Economy: consensus \$52.4bn, existing-business assumptions +\$17.8bn, agent revenue +\$53.4bn = \$123.5bn (25% of the gap comes from existing-business assumptions). Continuing spending is included in existing-business revenue in the published scenarios, so its additional bar is zero. Consensus is the FMP mean at the edition's snapshot, extended past 2029 using our assumed slowdown in growth. Modeled.

Is activity becoming paid use?

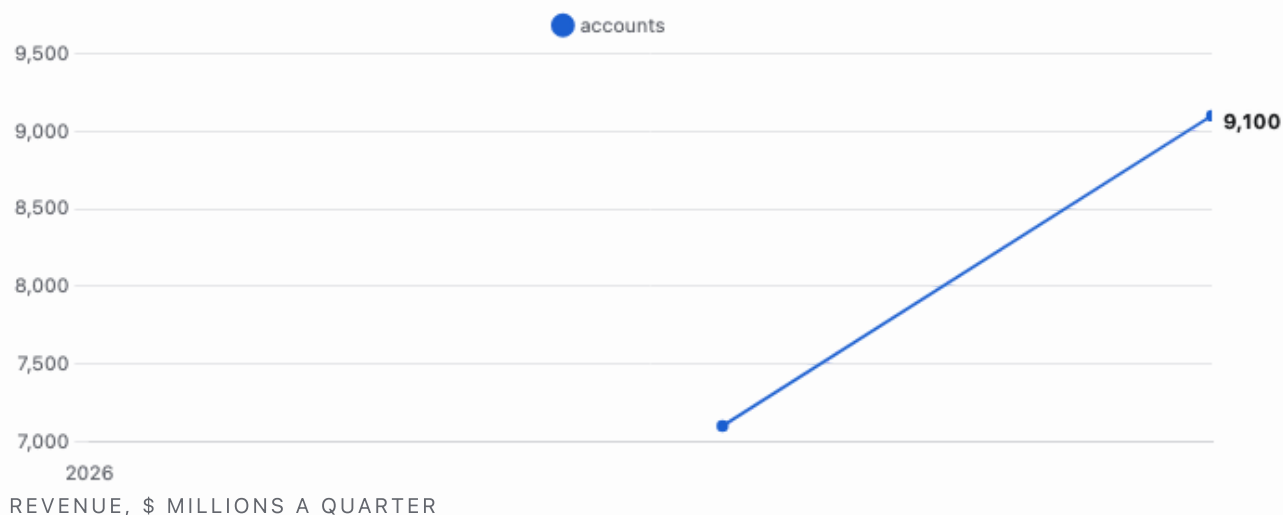


MONGODB · AI-NATIVE CUSTOMERS ON ATLAS · CUSTOMERS



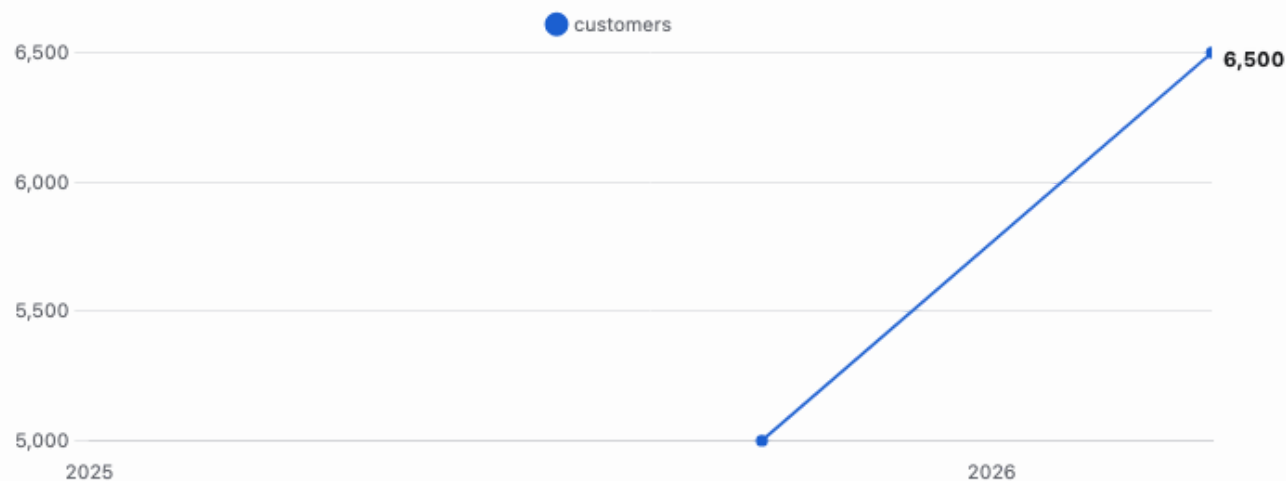
Reported regularly: the share of Atlas ARR from customers with an AI use case, customer revenue rather than AI-specific consumption. Revenue from AI-native customers grew 67%, 129% and 165% year over year across the three Q2s, from a low-single-digit share of revenue. Latest activity: 2,200+ (Q2 FY27) (source checked).

SNOWFLAKE · ACCOUNTS USING CORTEX CODE · ACCOUNTS



Disclosed once: a \$100 million AI revenue run rate on December 3, 2025 (Q3 FY26 call); no recurring standalone AI revenue series since. Management says Cortex Code accounts 'are consuming more of the core' (September 2, 2026); the consumption shows up in product revenue, not as a line. Latest activity: 9,100 accounts using Cortex Code; 2,000+ net new in the quarter (Q2 FY27) (source checked).

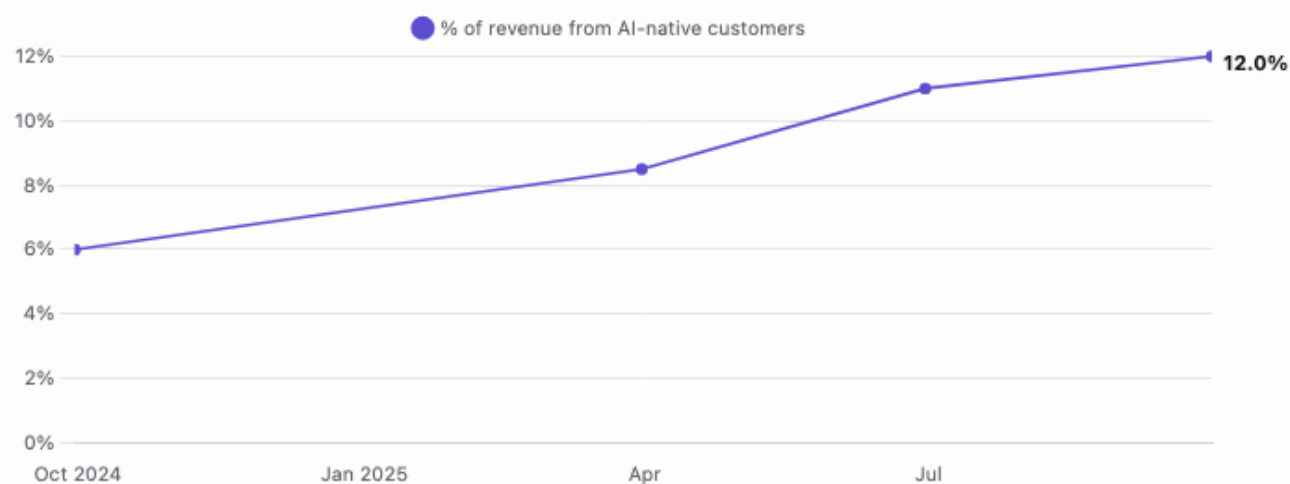
DATADOG · CUSTOMERS SENDING DATA THROUGH AI INTEGRATIONS · CUSTOMERS



REVENUE, \$ MILLIONS A QUARTER



% OF REVENUE FROM AI-NATIVE CUSTOMERS

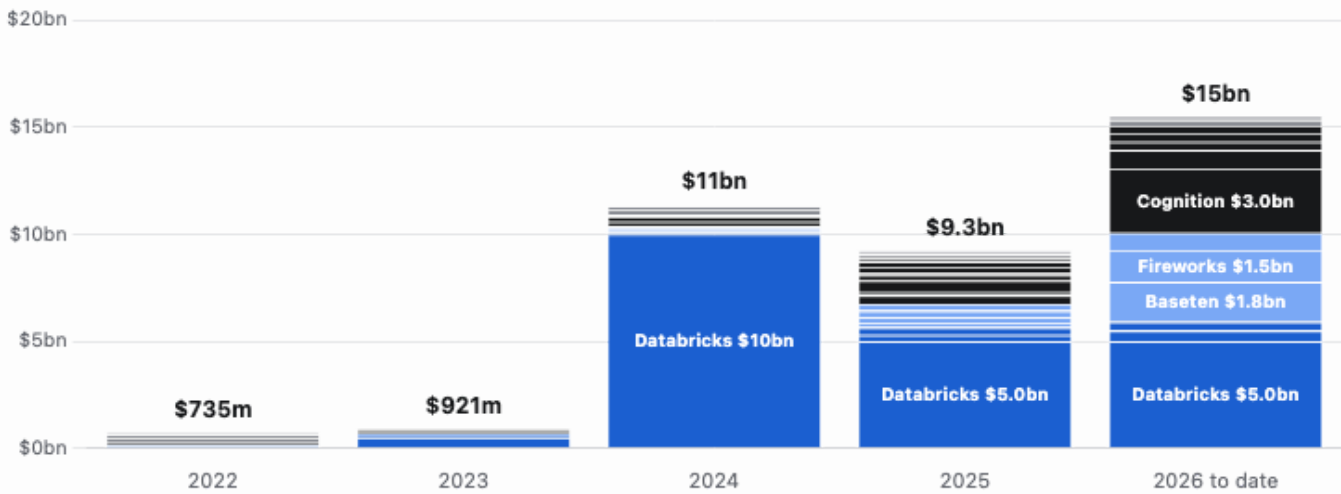


Reported through Q3 2025, then counts only: AI-native customers' share of revenue, customer revenue rather than AI-specific consumption. Datadog stopped disclosing the share after Q3 2025 and gives counts since: 650 AI-native customers in Q4 2025, 750+ in Q2 2026, 31 spending over \$1M a year. On August 6, 2026, Datadog reported lower usage from its largest customer and included it in Q3 and full-year guidance. Q3 revenue growth was guided to 28% to 29%. Latest activity: 6,500 customers, 20% of customers and 80% of ARR (Q1 2026) (source checked).

Inform: the near-term claim, that paid growth lasts, and whether agent activity reaches the meters. Activity counts, revenue and AI-attributable shares are different measures with different denominators; where a company discloses nothing the gap is left. Each point is dated at the period it describes and sourced in the record (monetization.json). Reported, except where marked derived.

2022 TO 2026 · UPDATED MONTHLY

Capital raised by the private backend



Primary equity rounds of 27 of the 28 private companies in Part Five’s map (the rest disclose no qualifying round), stacked by company and colored by layer: ● State, ● Inference cloud, ● Runtime, ● Control plane. Secondaries, debt and undisclosed amounts excluded. Record as of October 5, 2026; we refresh it monthly from SEC Form D. This is [Figure 5.3](#) in Part Five.

A3

Valuation

DAILY

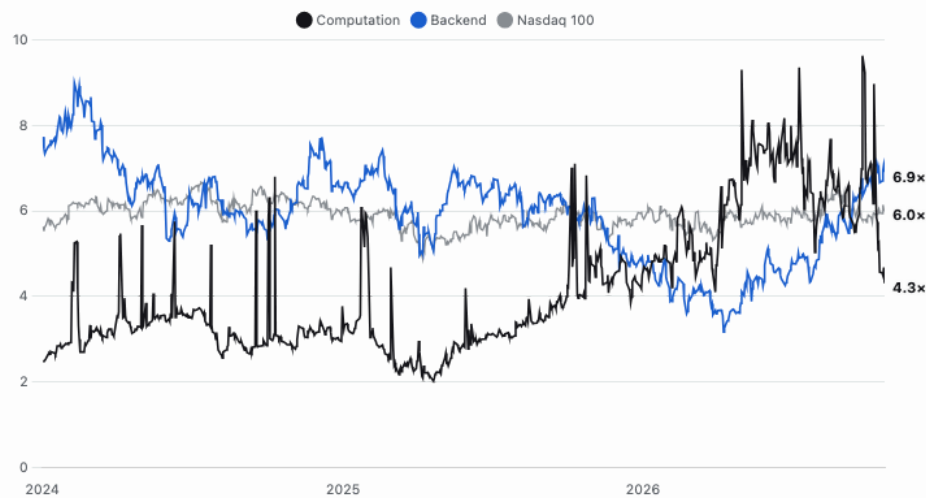
Figures 4.3 · 4.4 in the essay →

as of October 22, 2026 · forward figures use the latest consensus

VALUATION UNDER STATED ASSUMPTIONS

The backend trades at 8.6× trailing and 6.9× forward revenue, 19% above the Nasdaq 100 on trailing revenue. Under the stated cash-flow assumptions, its forward multiple corresponds to 16% annual growth for five years. For context, backend consensus growth to fiscal 2029 is about 23% a year. Coverage and periods differ, so this comparison does not establish a valuation gap. Computation trades at 10.7× trailing. As of October 22, 2026.

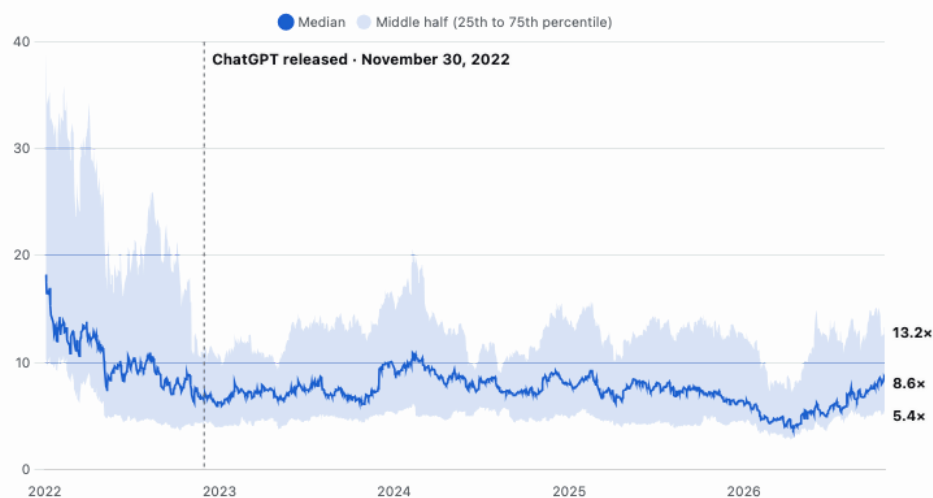
Forward revenue multiples



Median enterprise value to the next twelve months of revenue, daily, on the latest consensus; names without an estimate drop out. Latest: October 22, 2026 · backend 6.9x · Computation 4.3x · Nasdaq 100 6.0x. [Figure 4.3](#) in Part Four carries the trailing multiples.

Inside the backend

CONVERGING The band is closing: the names trade closer together. Spread narrow by the backend's own history.



The line is the median of the 31 backend names with a quarterly series; the band is the middle half, a quarter of the names below it and a quarter above. A narrow band means the middle half trade at similar multiples. It does not establish that individual companies moved together. The companies occupying each quartile can change.

Latest: October 22, 2026 · median 8.6x · middle half 5.4x to 13.2x, top quarter at 2.4 times the bottom. A year earlier 2.9 times: the backend has closed ranks.

Growth consistent with the backend's valuation

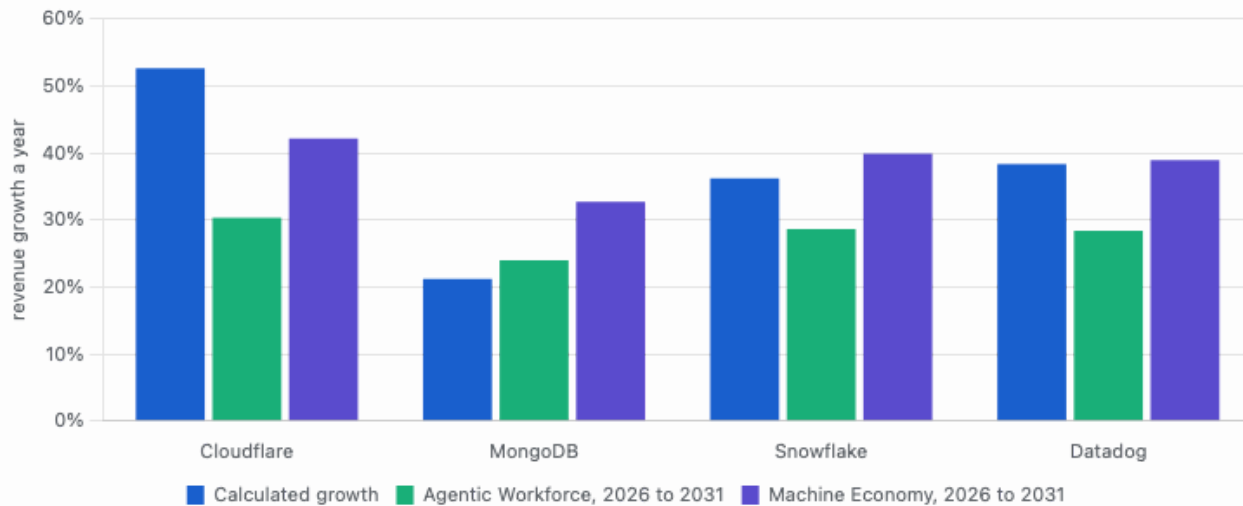
Different company groups and horizons; scale comparison only. Dotted lines are four-company scenarios, not forecasts for this basket.



The five-year growth rate that matches the backend's observed forward multiple under Figure 4.4's cash-flow assumptions: five years of growth fading to 3% by year 11, a 25% cash margin, a 10% discount rate. The multiple is the median of the 32-name backend universe, so the line is the growth consistent with the category's valuation under these assumptions, not a finding that the market is mispriced. For scale, the four company cases compound at 27.7% and 38.3% a year to 2032 on our two scenarios and 22.4% on consensus; the two scenario rates are the dotted lines, a different group on a different path, drawn for scale only. The card beside this one makes the like-for-like comparison, company by company, on one horizon, 2026 to 2031. Latest: October 22, 2026 · median forward multiple 6.9× corresponds to 16% a year · 31 of 31 with an estimate. This is [Figure 4.4](#), run daily.

COMPARED WITH ITS TWO SCENARIO PATHS

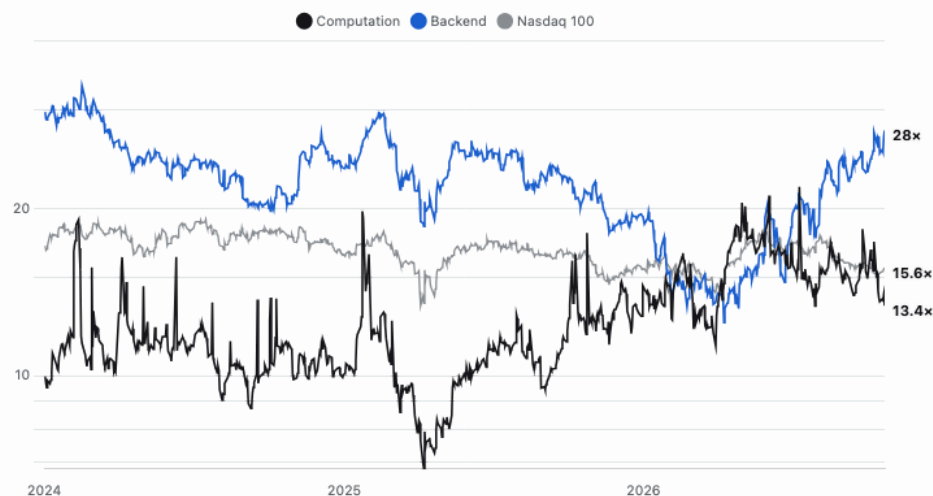
Growth consistent with each company's valuation



Like for like: each company's implied five-year growth (Figure 4.4's terms, growth applied from the first year because the multiple is trailing) beside the growth our two scenarios carry for that company from 2026 to 2031, the five years from the valuation date, the horizon the DCF holds growth for. Cloudflare 45.8x trailing revenue corresponds to 53%; the scenarios carry 30% and 42%. MongoDB 9.9x trailing revenue corresponds to 21%; the scenarios carry 24% and 33%. Snowflake 21.1x trailing revenue corresponds to 36%; the scenarios carry 29% and 40%. Datadog 23.4x trailing revenue corresponds to 38%; the scenarios carry 28% and 39%. Enterprise values as of October 23, 2026 (FMP). The blue bars show valuation-implied growth; green and purple show the two scenario paths. A higher blue bar means the valuation requires more growth than that scenario supplies, on these assumptions. Reported, derived and modeled: the multiple is reported, the implied growth derived, the scenario growth modeled.

DAILY FROM 2024

Enterprise value / next-twelve-month EBITDA



Median enterprise value to the next twelve months of EBITDA. Names with negative forward EBITDA drop out; EBITDA ignores capital spending, which flatters the Computation basket. Names with an estimate: backend 31 of 31 · Computation 25 of 25 · Nasdaq 100 95 of 101. Latest: October 22, 2026 · backend 27.6x · Computation 13.4x · Nasdaq 100 15.6x.

Enterprise value is market value plus debt less cash and short-term investments.

These are the series we read, not numbers we forecast. **Tokens**: how much of one routing platform is agents at work, what they do, what it costs, and the machine hours that implies. **Other adoption evidence**: the outside series behind the model's inputs. **Capability**: how hard a task a machine completes, measured in the human time it would take (METR), not how long an agent runs. **Production**: how much software gets written. We grade the bold bets of Part Four against them.

B1 Tokens

DAILY Figures 2.6 · 2.7 in the essay →

AGENT-APP TOKENS RISING 17.0 trillion a day, 73% of OpenRouter tokens, in the week ending October 18, 2026; 13.4 trillion four weeks earlier.

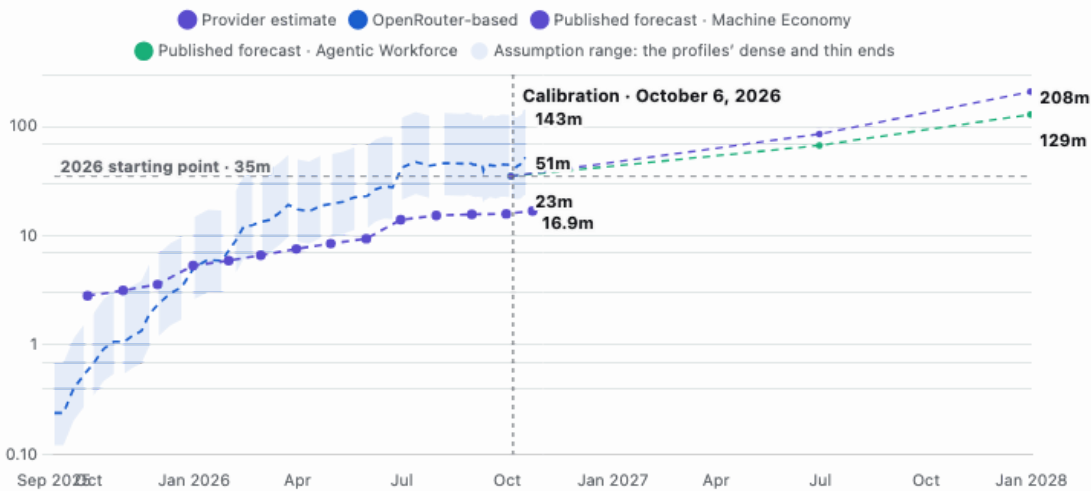
Estimated machine hours worldwide

55 million

machine hours a day, worldwide · week to October 18, 2026

Skycatcher estimate using OpenRouter data. We apply its agent share and workload mix to estimated worldwide token use. The separate provider-by-provider estimate is **17 million** a day. Neither is a direct count.

PROVIDER ESTIMATE MONTHLY; OPENROUTER ESTIMATE WEEKLY · SHADING VARIES THE ASSUMPTIONS · DASHED LINES MARK ESTIMATES, INTERPOLATIONS AND FORECASTS; NO WORLDWIDE HOURS ARE DIRECTLY OBSERVED · LOGARITHMIC SCALE



Week ending October 18, 2026: latest OpenRouter-based estimate, 55 million hours a day; alternative provider-by-provider estimate, 17 million. Original 2026 forecast starting point: 35 million a day. Skycatcher estimates, neither a direct count. The shaded range varies tokens used per active hour. It does not capture every uncertainty in global token use, the share doing agent work or whether OpenRouter's workload mix represents other platforms. It is not a confidence interval.

LATEST READING AGAINST THE REQUIRED PATH · MILLION MACHINE HOURS A DAY

55m

OpenRouter-based estimate, week to October 18, 2026 · 2.9x in six months, estimated across a method change

17m

provider-by-provider estimate, October 18, 2026

67 to 85m

needed by June 30, 2027, Agentic Workforce to Machine Economy

129 to 208m

needed by December 31, 2027

2.9x to 4.2x

growth a year the paths ask for, anchor to December 31, 2027

▼ Method and sources

Three assumptions determine the estimate. Their uncertainty is not combined into a confidence interval.

ASSUMPTION	EFFECT ON ESTIMATED HOURS	WHAT THE SHADED RANGE COVERS
Worldwide token volume	More global tokens imply more hours, holding the mix and token rates fixed.	Not included. Disclosures and estimates have coverage gaps.
Workload mix and agent classification	Changes which tokens count as agent work and how they are allocated.	Not included. Alternative allocations are shown separately; OpenRouter may differ from the wider market.
Tokens per active hour	More tokens per hour imply fewer hours for a fixed token volume.	Included: the dense and thin endpoints of the three workload profiles.

Week to October 18, 2026. Provider-by-provider estimate: 17 million machine hours a day across 12 venues, 2.0x in six months. OpenRouter-based estimate: OpenRouter’s 17.0 trillion agent tokens a day, split by tag into coding 49%, research 15% and workflow 35% (29% of tokens carry tags that do not say which workflow they belong to, allocated in proportion), each ÷ its profile’s tokens an hour (Appendix B1) and summed = 2.3 million on the platform; ÷ its 4.1% share of the world’s disclosed or estimated tokens = 55 million in the world. Allocating the uncertain tags all to coding gives 45 million, all to workflow 55, all to research 88; the profiles’ dense and thin ends give 25 to 154; 2.9x in six months across the August 7, 2026 method change; not a like-for-like observed growth rate. Activity converted to hours, a cross-check: 0.89 million, Devin (Cognition) 0.30 million as of September 25, 2026; Salesforce Agentforce 0.58 million as of July 31, 2026. The original 2026 forecast starting point is 35 million, set on October 6, 2026 on a one-operator coding profile (that basis reads 29 million this week) and kept until this calibration is reviewed under the rule in Appendix A4. The provider-by-provider and OpenRouter-based estimates are alternative methods for one quantity, not independent confirmations and not a confidence interval; OpenRouter is counted once in the provider estimate (4.1% of the world’s tokens). Fresh tokens the hours imply: 1.3 million an hour across the mix. What moved the OpenRouter estimate since September 20, 2026, as factors that multiply: ×1.232 in all = agent tokens on the platform ×1.273 × the adjustment for the platform’s share of the world’s tokens ×0.780 × calibration and mix ×1.240. Platform volume +27.3% while the OpenRouter estimate moved +23.2%: the platform grew faster than the world estimate it is scaled by. Both estimates are inferred from observed tokens: the provider estimate from each venue’s disclosed volume × an estimated agent share ÷ a profile’s tokens an hour, the estimate from OpenRouter’s task tags, which stand in for the composition of agent work and are not a classification of every agent token.

Allocated tokens and inferred hours · central assumptions		
WORKLOAD	TOKEN SHARE	HOURLY SHARE
CODING AGENTS	49.2%	18.4%
RESEARCH AND BROWSING AGENTS	15.4%	46.3%
ENTERPRISE WORKFLOW AGENTS	35.3%	35.3%

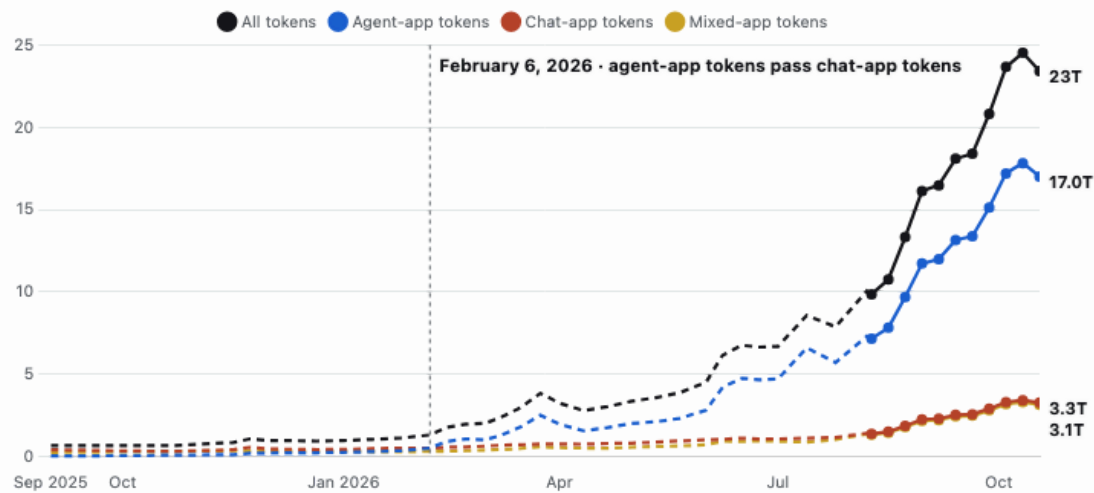
Hour share is each workload’s allocated tokens divided by its assumed tokens per hour, as a share of total inferred hours. These are model contributions, not observed shares of worldwide agent work.

No worldwide count exists. We estimate the same quantity in two ways. The **provider-by-provider estimate** combines platforms: daily tokens × estimated agent share ÷ tokens per agent hour. We remove routed traffic from the provider totals so it is counted once. It updates as new disclosures arrive. Workload assumptions are in Appendix B1. The **OpenRouter-based estimate** starts with OpenRouter’s agent tokens. We convert coding, research and workflow tokens to hours, then divide by OpenRouter’s estimated share of worldwide token use. Unclear task tags

follow a stated allocation rule, with alternatives shown. This method assumes one platform's mix helps describe the wider market. The shading varies token use per hour across the three workload profiles. It does not include all uncertainty about worldwide token volume, classification or whether OpenRouter's mix represents the wider market. **Activity converted to hours**, shown in the reading, provides another comparison from products that bill hours or actions. It is not a measured lower bound. Both main estimates share workload assumptions, so agreement is not independent confirmation and their range is not a confidence interval. The dashed lines show the **published forecast**: a fixed starting point of 35 million hours a day, set on October 6, 2026, and each scenario's path through 2027. The marked method changes are explained in Appendix A4. Estimates may miss uncovered platforms and tool work that uses few tokens.

AGENT, CHAT AND MIXED APPS · TRILLION A DAY · 7-DAY AVERAGE

Token use on OpenRouter



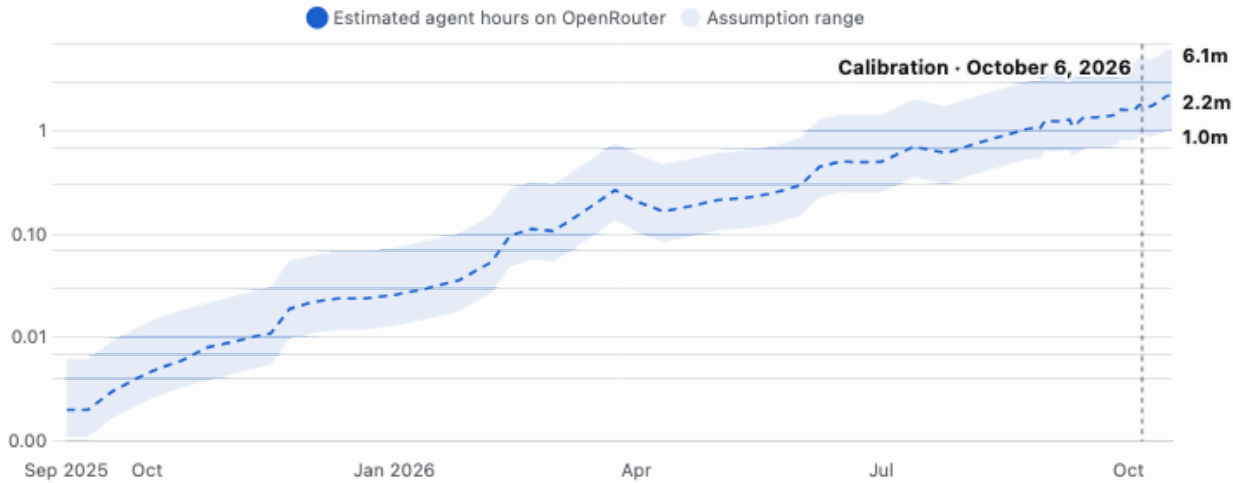
Week to October 18, 2026: 23.4 trillion tokens a day; agent apps 17.0 trillion (73%), chat apps 3.3, mixed apps 3.1 · about 3.9% of the world's disclosed and estimated tokens · [OpenRouter rankings](#)

Total platform traffic, split by application type. Dashed history through August 7, 2026 was read from OpenRouter's published chart; later dots use weekly volumes and app classification. Lines between dots are guides, not daily observations. The marked method change matters when comparing periods. This is one platform's traffic.

MILLION HOURS A DAY · ASSUMPTION RANGE SHADED

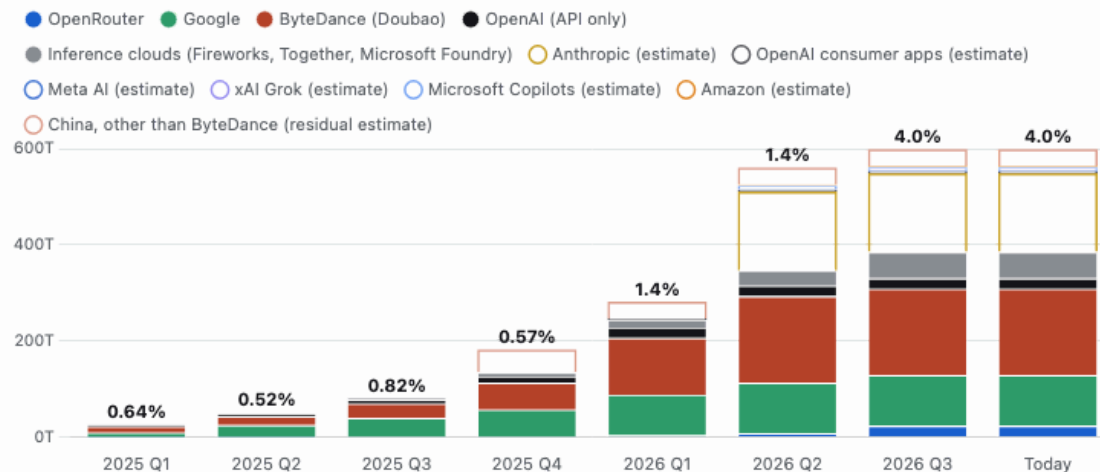
Estimated agent hours on OpenRouter

TOKENS FROM THE CHART ABOVE CONVERTED TO HOURS · WORKLOAD ASSUMPTIONS SHADED · LOGARITHMIC SCALE



Agent tokens the platform processed, split by tag and divided by each workload's tokens an hour (Appendix B1), summed: 2.3 million machine hours a day in the week to October 18, 2026, 1.0 to 6.3 across the workload assumptions. The tokens themselves are observed (the chart above); the hours are inferred from them. Earlier conversion bases, the one-operator cadence and the separate fresh basis, are described in Appendix A4 and A2.

OpenRouter against the world's tokens



Today: about 599 trillion tokens a day disclosed or estimated across the world (about 385 trillion disclosed, 214 trillion estimated for gaps in disclosures), of which OpenRouter carries 23.7 trillion, a 4.0% share. Annualized that is about 219 quadrillion a year. Anthropic is 77% of the estimated layer, from a range of 60 to 164 trillion; at the low end the world is about 495 trillion and OpenRouter's share 4.8%. Estimates begin in 2025 Q4; earlier bars are official only. The proxy below is rebased to this share while the estimates are shown.

Both views interpolate between dated statements and hold the last value until a new disclosure. The default view includes estimated providers; the other view includes disclosed providers only. Quarterly bars are estimates at quarter-end, not direct measurements of every date. Coverage is incomplete: several major providers publish no volume, and OpenAI's API disclosure excludes consumer ChatGPT. Our estimates fill gaps where providers publish no volume and where the China residual cannot be reconciled from the disclosed inputs. Estimated values appear as outlines and begin in the stated quarters. OpenRouter's share changes with the selected total. The source record and calculation are under Method.

► [Method and the dated record](#)

▼ How the OpenRouter estimate is built · agent-hour measurements and tokens per request

OPENROUTER'S GLOBAL SHARE AND WORKLOAD MIX

Assumptions

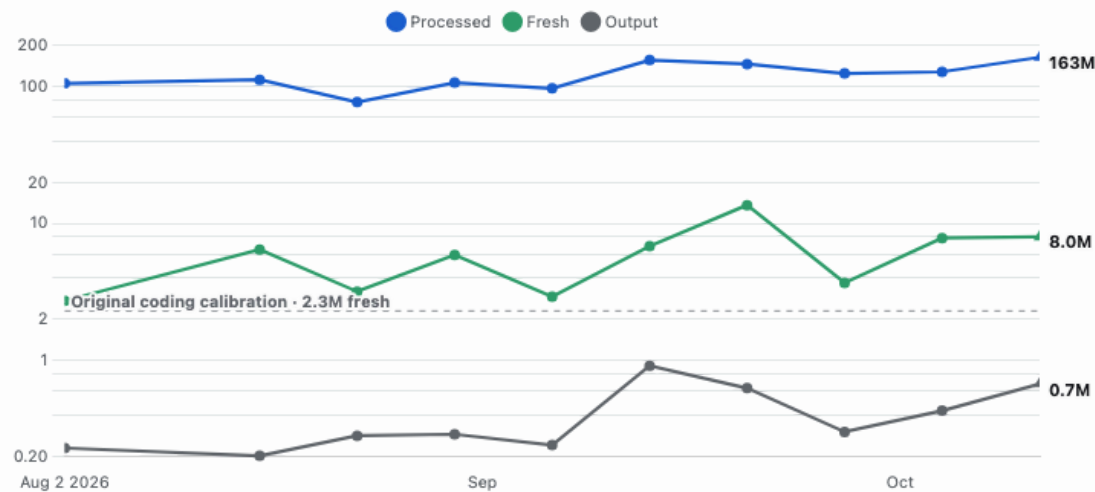
OpenRouter's share of the world's tokens **4.1%**

Each reading retains its task mix from October 7, 2026 onward. Earlier readings use the first recorded mix. OpenRouter's sampled task shares and our workload allocation use different denominators. Task window ending October 17, 2026): coding 35%, research 11%, enterprise workflow 25%, workload uncertain 29% (planning, tool dispatch, memory, classification, data extraction). Estimated composition of agent workloads after allocating the uncertain share in proportion: coding 49%, research 15%, enterprise workflow 35%. The gauge quotes the hours under the three other allocations as sensitivities.

Worldwide hours equal estimated platform hours divided by the platform's global share. Halving that share doubles the result. The starting share includes our estimates for providers without disclosures. OpenRouter classifies sampled traffic by task; we allocate those tasks to workload profiles. The two breakdowns use different denominators.

PROCESSED, FRESH INPUT AND OUTPUT TOKENS

What an active agent hour uses



The weeks to October 7, 2026: medians of 257 model calls an hour, 418k tokens of context a call, 112 million tokens processed an hour (95% cache reads), 5.9 million fresh and 289k written (execution-interval basis since October 7, 2026: hours while the model generates or a tool runs, waiting on a person excluded). The model converts hours to tokens at 1.3 million fresh tokens a machine hour for 2026, the blended workload profile refreshed with the data; the 2.3 million the anchor was set on is this operator's October 6 figure. A triangle marks the partial week ending October 11, 2026. Medians use complete weeks only.

Our session logs count intervals when the model generates or a tool runs, capped at ten minutes. Waiting for a person is excluded; a parent's wait for a child counts only in the child's work. Processed tokens include everything read and written. Fresh input includes new input and cache writes; output is reported separately. This is one operator's sample. Appendix A2 preserves the earlier capped-gap estimate.

AGENT APPS · TRAILING 30 DAYS

Tokens per request on OpenRouter

Thirty days to October 22, 2026: the platform's agent apps carried about 108k tokens a request, token-weighted across 14 apps (84% of the top fifty's tokens), quartiles 92k to 138k; the platform as a whole 82k. At our agents' 257 calls an hour a platform agent hour processes about 28.0 million tokens on the one-operator profile; the OpenRouter-based estimate uses the workload blend of 7.5 million.

The record began on October 6, 2026; the chart of tokens a request over time draws once it is thirty days long. Until then the latest reading is above, and the app table is in Appendix A4.

OpenRouter publishes tokens and requests by app over the preceding thirty days. Dividing them estimates tokens processed per request. Comparing this with fresh input per call in our logs helps estimate how much context is being reused. App classification is in Appendix A4; the limits of that comparison are in Appendix A2.

What the tokens are used for

KIND OF WORK	FIRST · SEPTEMBER 27, 2026	LATEST · OCTOBER 22, 2026
Agent work	35.9%	35.3%
Code	36.6%	29.4%
Data	5.9%	6.8%
General use	21.5%	28.6%

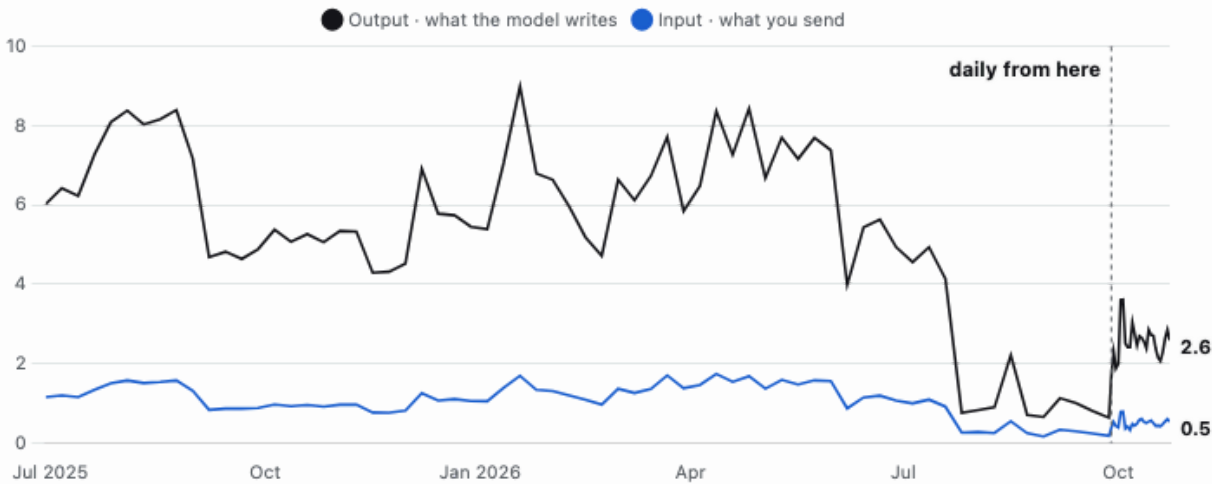
Drawn as a table until the record has 30 daily readings.

Latest window ending October 22, 2026 · Agent work 35.3% · Code 29.4% · Data 6.8% · General use 28.6% · largest tasks: Workflow Execution 23.7%, Classification 10.1%, Code Generation 9.2%, Multi-step Planning 8.4%, Debugging 5.5%.

OpenRouter’s task labels are grouped into agent work, code, data and general use. Each point covers seven days. The record begins September 27, 2026. This task-based definition of agent work is narrower than the app-based classification above, so the shares differ.

What tokens cost

Price of a million tokens



Latest measured October 23, 2026 · a million input tokens \$0.54 and output \$2.58. Week to June 29, 2025: \$1.16 in, \$6.02 out.

Input is what the model reads; output is what it writes. The chart weights OpenRouter list prices by model token use. Weekly observations start in June 2025, with daily observations from September 28, 2026 covering the twenty busiest paid models. The join changes the basket as well as the frequency; it should not be read as a pure price change.



BY TURN COUNT · DOLLARS · DAILY SINCE SEPTEMBER 28, 2026 · OPEN TO SEE

What a Claude Code session costs

Turns in the session	First · September 28, 2026	Latest · October 23, 2026
1 turn	\$0.017	\$0.018
2 to 9 turns	\$0.034	\$0.044
10 to 49 turns	\$0.600	\$0.859

Drawn as a table until the record has 30 daily readings.

Latest snapshot October 23, 2026, the 26th daily reading since September 28, 2026 · 1 turn: \$0.018 (median across 33 models \$0.010), about 12k tokens (middle half 7k to 17k), so \$1.51 per million tokens · 2 to 9 turns: \$0.044 (median across 45 models \$0.035), about 35k tokens (middle half 24k to 42k), so \$1.25 per million tokens · 10 to 49 turns: \$0.859 (median across 22 models \$0.815), about 199k tokens (middle half 160k to 229k), so \$4.33 per million tokens.

A session is one run; a turn is one exchange with the model. The lines group sessions by turn count and weight OpenRouter’s median costs across the models used. Each point is a daily snapshot.

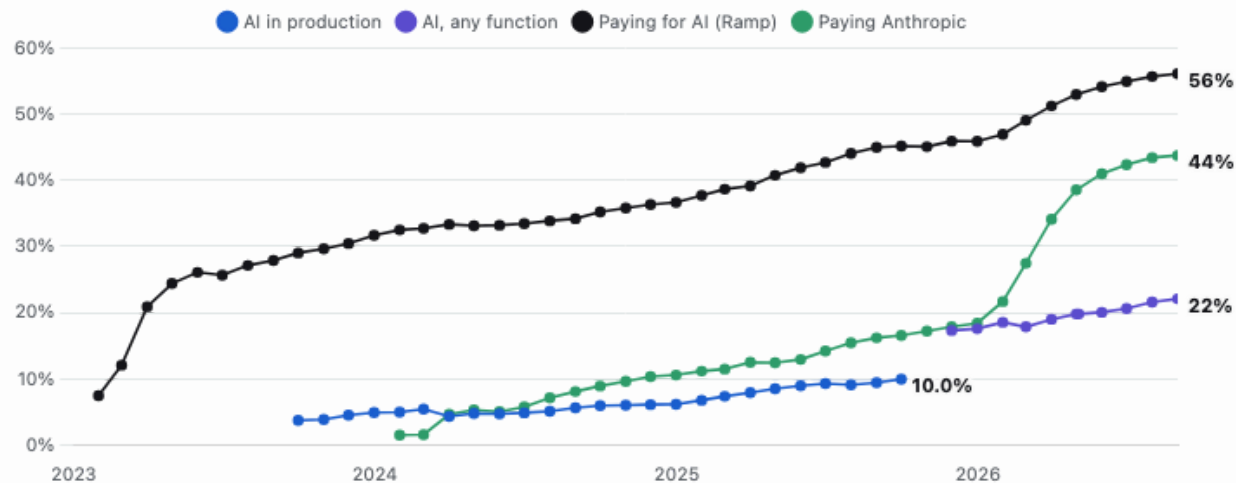
ADOPTION 22% OF FIRMS · 32% OF WORKERS WEEKLY

Census: 22.1% of US firms used AI in any business function in the two weeks to August 31, 2026. Gallup: 16% of US employees use AI daily and 32% a few times a week or more (August 31, 2026). Ramp: 56.1% of businesses in Ramp’s dataset paid for AI in August 2026. The model assumes 12% of knowledge workers worldwide use agents in 2026. Gallup measures broader AI use among US employees, so it provides context rather than a direct estimate of agent adoption.

These sources help assess the model’s inputs: business and employee adoption, spending, autonomy, agent activity, tool use and web traffic. Each observation retains its source and date. Comparable readings form charts; one-off disclosures remain in the tables in [Appendix B4](#), and [Appendix B1](#) says which input each series grades. Refreshed monthly, Cloudflare Radar weekly.

Adoption at the firm

PERCENT OF US FIRMS · THE CENSUS QUESTION CHANGED IN NOVEMBER 2025, SO ITS LINE BREAKS; RAMP COUNTS PAID USE, AND THE SHARE PAYING ANTHROPIC



Inform: Share of knowledge workers who use agents (the model holds 12% for 2026; the 2032 destinations are 70% and 95%); the Anthropic line is the agent-heavy vendor's reach, a check on the bridge's Anthropic venue

AI in production · biweekly; monthly means shown · latest Census BTOS national AI use rate, monthly mean of 2 biweekly reading(s): 9.95% (earlier Census question) (September 30, 2025, reported via secondary source, ramp.com).

AI, any function · biweekly; monthly means shown · latest Census BTOS national AI use rate, monthly mean of 2 biweekly reading(s): 22.1% (broader question introduced November 2025); Monthly mean of two biweekly readings: 21.8% and 22.4%. (August 31, 2026, reported via secondary source, ramp.com).

Paying for AI (Ramp) · monthly · latest Ramp AI Index adoption rate 56.13% of businesses in Ramp's card-spend dataset paying for AI (interactive series as of 2026-10-06) (August 31, 2026, source checked, ramp.com).

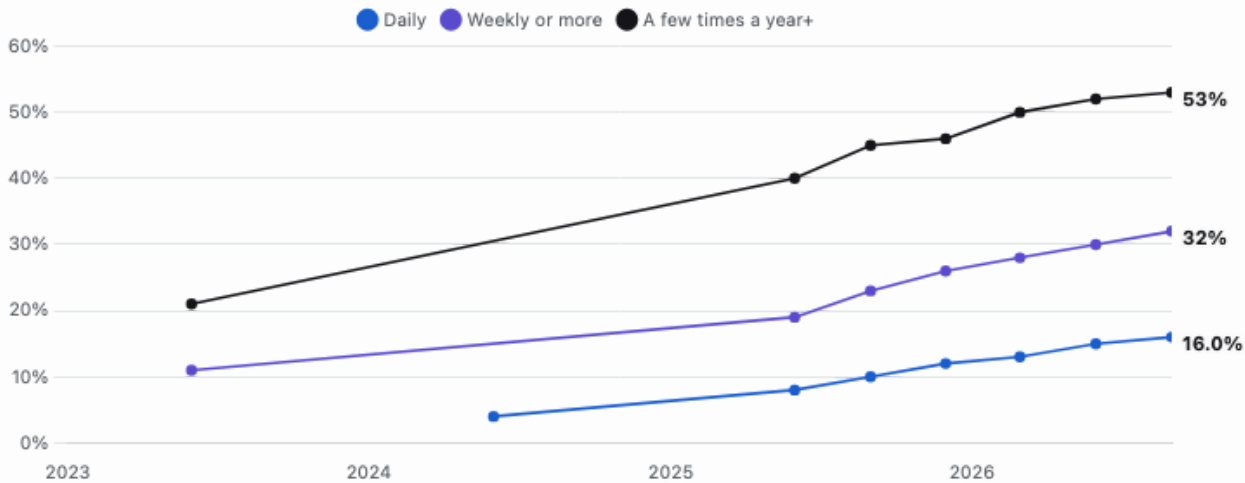
Paying Anthropic · monthly · latest Anthropic: 43.78% of US businesses on Ramp paying for Anthropic subscriptions or tokens (interactive series as of 2026-10-06) (August 31, 2026, source checked, ramp.com).

► Every point, its definition and its source

GALLUP

Adoption at the worker

PERCENT OF US EMPLOYEES · DAILY, A FEW TIMES A WEEK OR MORE, A FEW TIMES A YEAR OR MORE



Inform: Context for the 12% agent-adoption assumption in 2026. Gallup measures broader AI use among US employees, not agent use among knowledge workers worldwide; it does not directly establish this input.

Daily · quarterly · latest 16% use AI at work daily (August 2026 data-table row) (August 31, 2026, source checked, [gallup.com](https://www.gallup.com)).

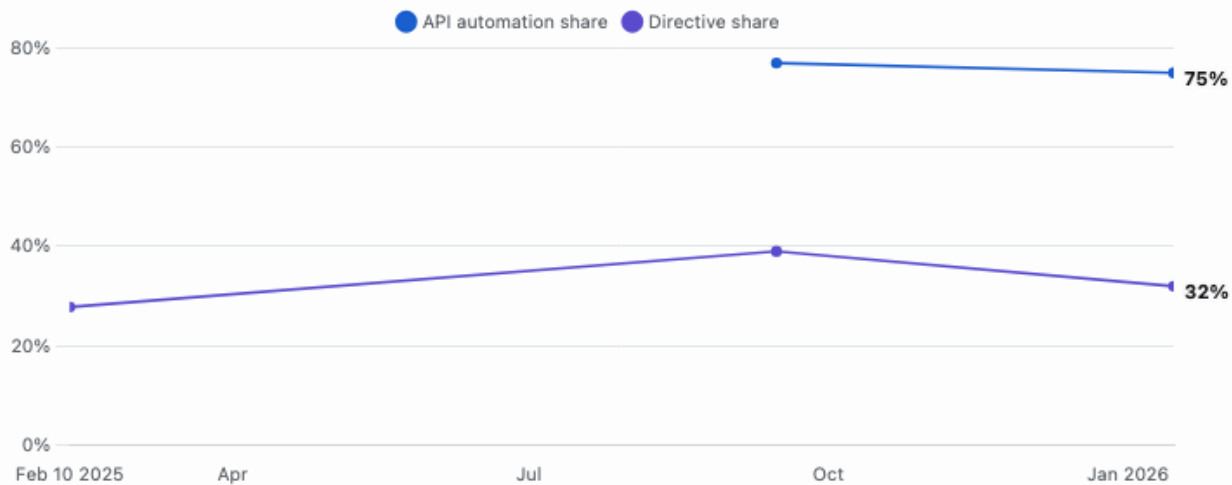
Weekly or more · quarterly · latest 32% use AI at work a few times a week or more (August 2026 data-table row) (August 31, 2026, source checked, [gallup.com](https://www.gallup.com)).

A few times a year+ · quarterly · latest 53% use AI at work a few times a year or more (August 2026 data-table row) (August 31, 2026, source checked, [gallup.com](https://www.gallup.com)).

► Every point, its definition and its source

Autonomy

PERCENT · AUTOMATION SHARE OF API USE; DIRECTIVE SHARE OF CLAUDE.AI CONVERSATIONS



Inform: Autonomous share of agents and hours per agent; the thesis's main assumptions, with no direct global measure yet

API automation share · per report, roughly quarterly · latest API usage is overwhelmingly work-related (74% vs. 46%) and directive (64% vs. 32%), with three-quarters of interactions classified as automation compared to less than half on Claude.ai (data Nov 13-20, 2025) (January 15, 2026, source checked, anthropic.com).

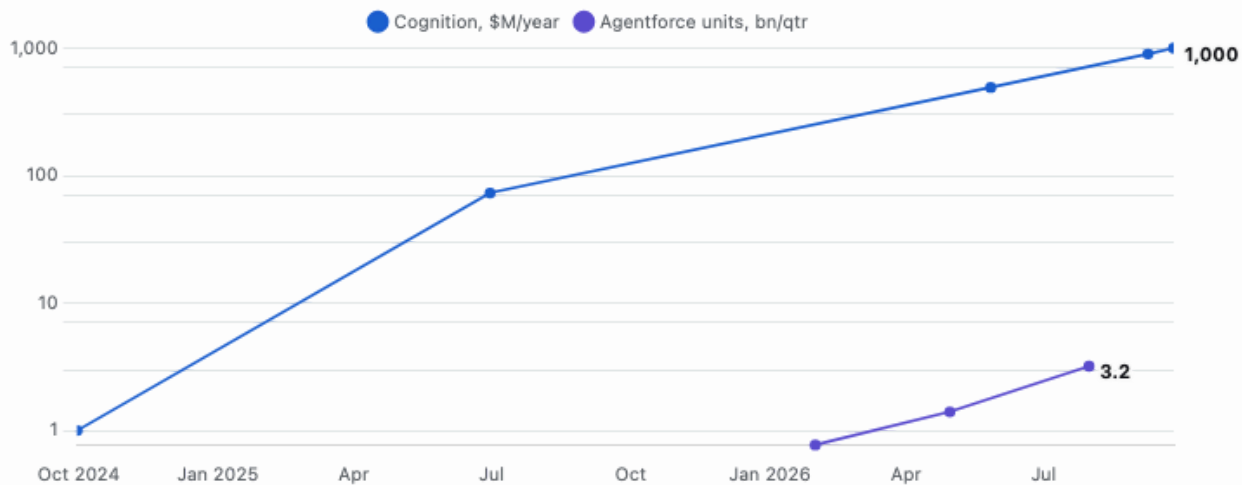
Directive share · per report · latest the share of directive conversations had fallen 7pp to 32% in November 2025; augmented jumped 5pp to 52% and automated fell 4pp to 45% (data Nov 13-20, 2025) (January 15, 2026, source checked, anthropic.com).

► [Every point, its definition and its source](#)

COGNITION REVENUE AND AGENTFORCE WORK UNITS

Activity proxies

COGNITION IN \$ MILLIONS A YEAR; AGENTFORCE IN BILLION WORK UNITS A QUARTER · LOG SCALE



Inform: Two different activity proxies. Cognition is company revenue, including Windsurf IDE seats from July 2025; it does not directly measure Devin hours. Any conversion to hours is illustrative.

Cognition, \$M/year · as disclosed · latest Cognition Crosses \$1B in Annualized Revenue Run Rate (September 25, 2026, source checked, cognition.com).

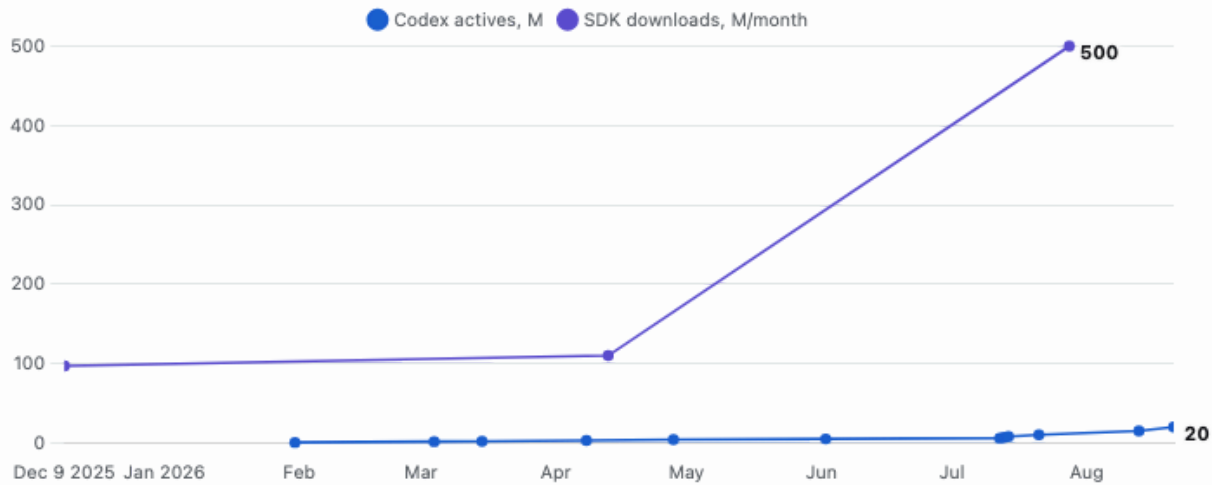
Agentforce units, bn/qtr · quarterly, earnings releases · latest 3.2 billion in Q2, growing 97% quarter-over-quarter (Q2 FY27, quarter ended 2026-07-31) (July 31, 2026, source checked, investor.salesforce.com).

► [Every point, its definition and its source](#)

CODEX USERS AND MCP DOWNLOADS

Agent users and software connections

MILLIONS · CODEX WEEKLY ACTIVES TO JUNE 2026, A COMBINED COUNT AFTER; MCP SDK DOWNLOADS A MONTH



Inform: Instructed agents per user and diffusion from the frontier; the tool-call plumbing agents run on, growth only

Codex actives, M · as disclosed · latest we have hit 20M active users for Codex some time this week (Tibo Sottiaux, X); memeburn: 'OpenAI Codex and ChatGPT Work have crossed 20 million active users' (August 21, 2026, reported via secondary source, explainx.ai).

SDK downloads, M/month · as disclosed · latest Across our Tier 1 SDKs, we're seeing close to half-a-billion downloads a month, with both TypeScript and Python SDKs crossing the 1 billion total downloads threshold. (July 28, 2026, source checked, blog.modelcontextprotocol.io).

► Every point, its definition and its source

Tokens used per active service-agent hour

OUR OWN TOOL-CALLING AGENT ON TAU-BENCH TASKS, MONTHLY · MILLION TOKENS AN HOUR OF AGENT EXECUTION, PROCESSED AND FRESH · LOG SCALE

SERIES	LATEST	DATE	BASIS
Processed, M an hour	13.0M an hour, 48 tasks, Claude Haiku 4.5, cache share 89%, pass rate 52%	October 7, 2026	measured by Skycatcher
Fresh, M an hour	1.07M fresh input an hour (0.32M written), 48 tasks, Claude Haiku 4.5, cache share 89%, pass rate 52%	October 7, 2026	measured by Skycatcher

The chart draws once a series has three readings; the record below holds every point.

Inform: The enterprise workflow profile of the hours gauge (B1), and the trigger in A4 for moving the anchor. Useful work from the same trace, October 7, 2026: 52% of 48 tasks passed, about \$0.05 of list-price inference a passed task, about 21 seconds of agent time a task; hours measure time, these measure use.

Processed, M an hour · monthly, by the measurement workflow · latest 13.0M an hour, 48 tasks, Claude Haiku 4.5, cache share 89%, pass rate 52% (October 7, 2026, measured by Skycatcher, scripts/measure-workflow-agent.py).

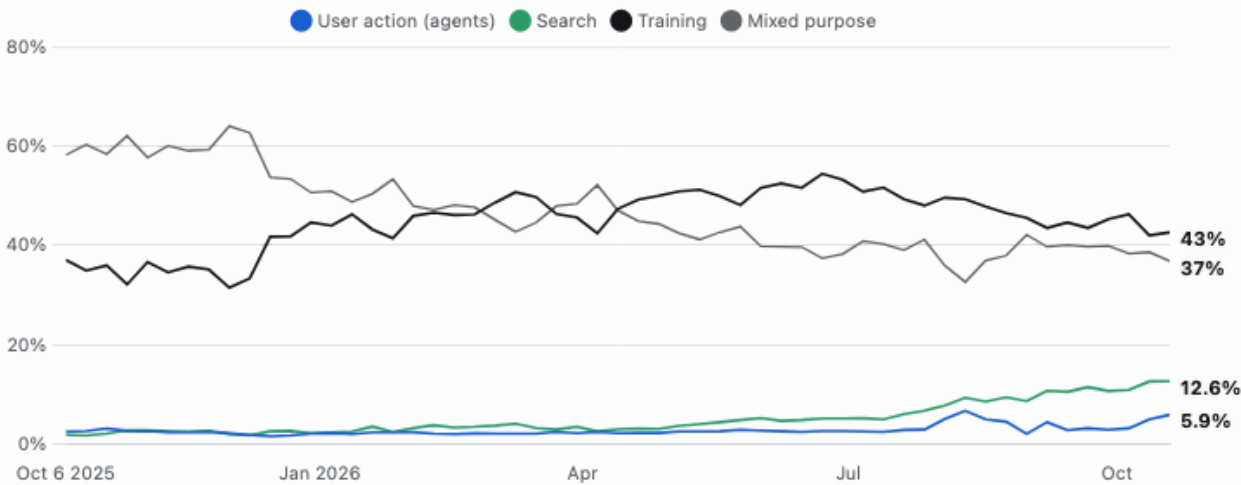
Fresh, M an hour · monthly, by the measurement workflow · latest 1.07M fresh input an hour (0.32M written), 48 tasks, Claude Haiku 4.5, cache share 89%, pass rate 52% (October 7, 2026, measured by Skycatcher, scripts/measure-workflow-agent.py).

► [Every point, its definition and its source](#)

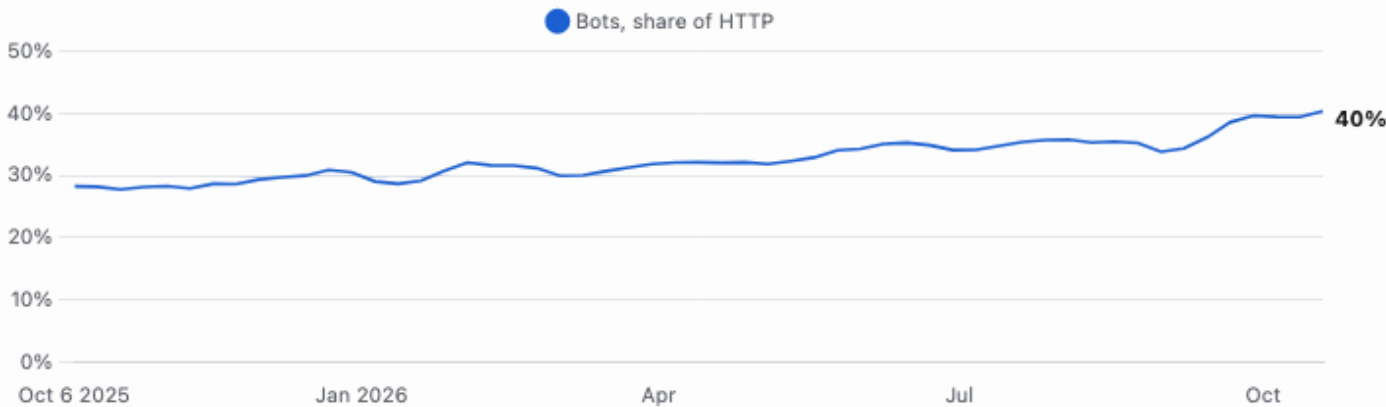
WHAT AI BOTS DO ON THE WEB · WEEKLY SHARES ACROSS ABOUT A FIFTH OF THE WEB · SINCE OCTOBER 6, 2025

Cloudflare Radar

SHARE OF AI-BOT REQUESTS BY DECLARED PURPOSE · PERCENT · USER ACTION IS AN AGENT FETCHING A PAGE ON A PERSON’S BEHALF



BOTS’ SHARE OF ALL HTTP REQUESTS ON CLOUDFLARE’S NETWORK · PERCENT · WEEKLY



Week of October 19, 2026: user action 5.9% of AI-bot requests (from 2.4% in the week of October 6, 2025), search 12.6%, training 42.6%; bots 40.3% of all HTTP requests (from 28.3%). Updated October 20, 2026.

AI BOT, WEEK OF OCTOBER 19, 2026	SHARE OF AI-BOT REQUESTS
Googlebot	22.4%
ClaudeBot	13.4%
Meta-ExternalAgent	7.4%
Bingbot	7.3%
Applebot	7.1%
GPTBot	5.9%
Amazonbot	4.6%
OAI-SearchBot	4.3%

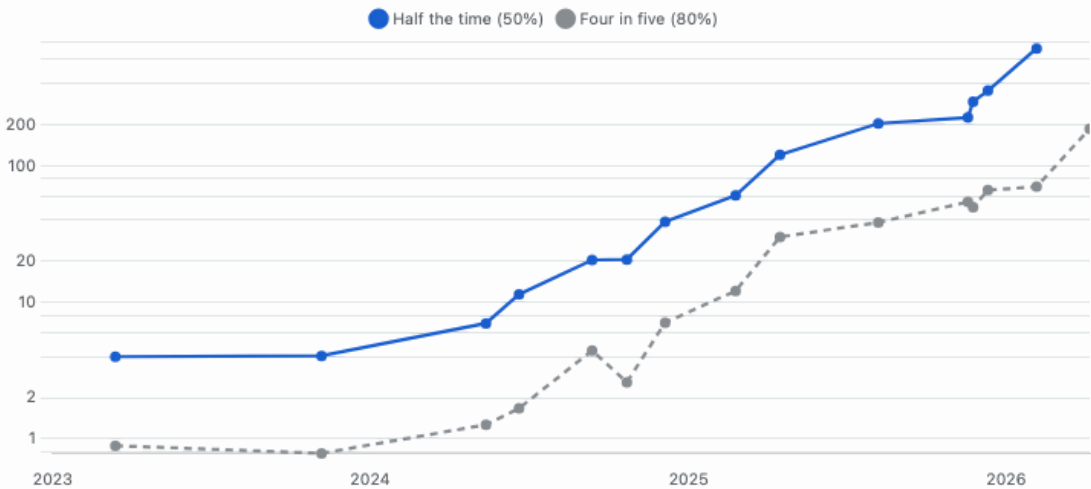
Informs: autonomous and browsing agents’ actions on the web, another measure beside OpenRouter’s tokens. Shares across Cloudflare’s network, about a fifth of the web; Radar publishes no request counts. Source: Cloudflare Radar API, weekly. Measured.

► Definitions and the record’s discipline

B3 Capability

HISTORICAL METR SERIES Figure 2.4 in the essay →

HUMAN-EXPERT MINUTES · LOG SCALE
Task length the best public model completes



Historical METR Time Horizon 1.1 series. The [source page](#) is no longer actively updated (notice dated September 8, 2026). Model release dates are shown; October 23, 2026 is the fetch date, not a new evaluation date. METR expresses task difficulty as the time a human expert would need. The chart shows frontier models at 50% success on a logarithmic scale. It does not measure how long the agent ran. Estimates above sixteen expert hours are excluded because the task suite cannot support them. The table includes each evaluated model and its confidence interval.

► Every evaluated model and confidence interval

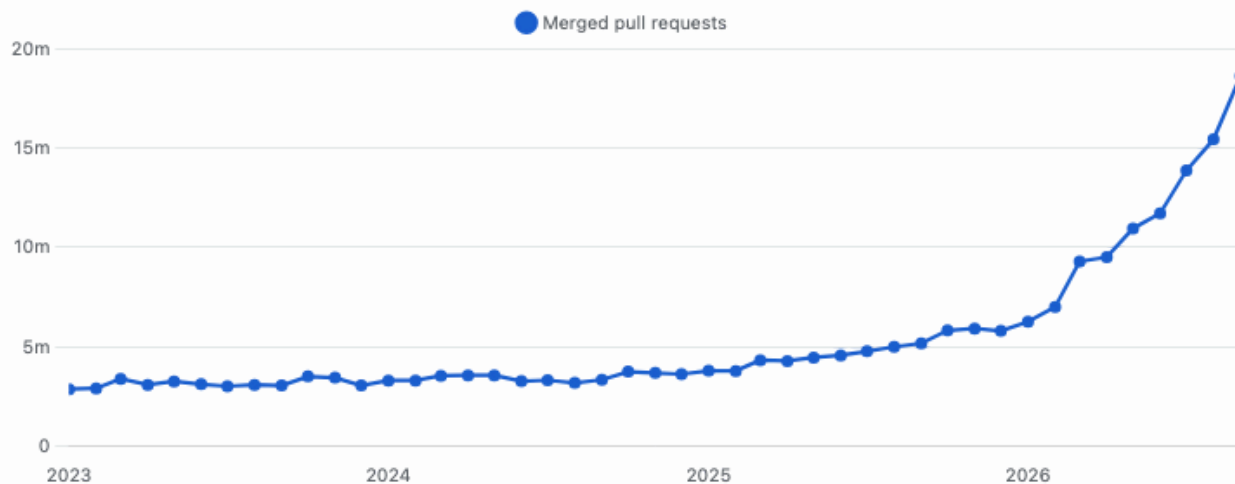
B4 Production

MONTHLY Figure 1.6 in the essay →

PUBLIC GITHUB PULL REQUESTS UP
261% YEAR OVER YEAR

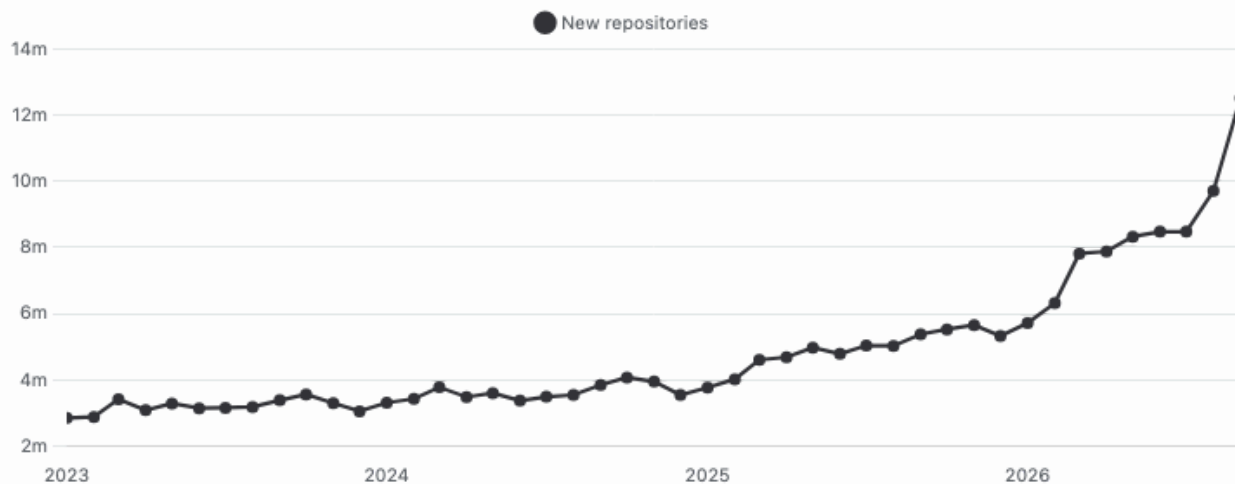
18.7 million pull requests merged on public GitHub in September 2026, 261% more than a year earlier and 6.5 times January 2023; new repositories 12.5 million. Pull requests have risen every month for the last four.

Merged pull requests a month



The share merged by coding agents is in [Appendix B3](#).

New repositories a month



Last closed month: September 2026.

Public repositories only; the running month is never drawn. The weaker commits series is in [Appendix B3](#).

MEASURES, DEADLINES AND RESULTS · LAUNCH BETS, OCTOBER 8, 2026

Dated predictions

THE HEADLINE PREDICTION · H1

Machine hours exceed all paid human work, 7.0 trillion hours a year, by the 2032 reading. Graded at the 2032 reading, by January 31, 2033.

▼ How it is measured and adjudicated

Observe the last complete week of December 2032; multiply daily active agent hours by 365. Compare with the fixed 2026 paid-human-work benchmark: central 7.0 trillion annual hours, range 6.5–7.5 trillion. A robust hit requires both the platform level and OpenRouter pulse lower sensitivity bounds above 7.5 trillion. A miss requires both upper bounds below 6.5 trillion. All overlap, disagreement, missing evidence or failed method review is unresolved. Publish the two central estimates as well as ranges; these are assumption ranges, not statistical confidence intervals. Later evidence can restate that week, never count activity after it. Initial grade January 31, 2033; one final restatement by June 30, 2033 may remain unresolved. An independent methodological review must be published by September 30, 2032. A replacement measured panel and its uncertainty rule may govern only if registered with that review at least a quarter in advance. Method changes are recorded under Appendix A4 and do not count as activity growth. The claim concerns active duration, not equivalent output or displaced jobs.

11

bets on the record

11

open

0

hit

0

missed

6

next 12 months · first grade June 30, 2027

5

beyond 12 months · first grade January 31, 2028

▼ Show all 11 bets, each with its measure and deadline

Next 12 months · by October 1, 2027

BET	DEADLINE	STATUS
B1 · The backend index beats the Computation layer over the 12 months from October 1, 2026.	October 1, 2027	Open
▼ How it is measured		
The equal-weight 32-name backend index against the 25-name Computation layer basket, both rebased to 0% at the September 30, 2026 close, the last close before October 1, 2026,		

BET	DEADLINE	STATUS
and read at the October 1, 2027 close on the tracker's index chart, start set to Since October 1, 2026. Both baskets are frozen at the launch roster; the roster rule below governs every later change. Launch roster rule: membership is frozen in launch-rosters.json. No new IPO or discretionary addition enters. Equal-weight daily returns use the same adjusted-close method as the published index. A delisting uses the last available adjusted close, then that position stays in cash at a zero return until the next trading day, when surviving members are equally reweighted; no replacement enters. If reliable delisting proceeds are unavailable, publish the missing treatment and leave the grade unresolved. Splits and dividends follow the adjusted-close series; acquisitions never add a new name. The launch-cohort grade must be published separately if the live tracker roster later changes.		
<p>B2 · Median backend revenue growth improves at least three points from the quarter ended June 30, 2026.</p> <p>▼ How it is measured</p> <p>Median year-over-year revenue growth of the matched panel: the backend companies that report a comparable quarterly figure for both the baseline quarter and the test quarter (at the baseline, 30 of 32; Megaport reports half-yearly and is excluded). Baseline 22.65% for the quarter ended June 30, 2026, so the threshold is 25.65%. The observation period is the quarter ended June 30, 2027 (or each company's last quarter ending on or before that date); the grade is read by September 30, 2027, allowing time for filings. Missing or incomparable test-period filings leave the bet unresolved; companies cannot be dropped after the fact to improve the median.</p>	September 30, 2027	Open
<p>B3 · At least 17 of the 32 backend companies disclose an aggregate AI-attributable revenue, ARR, run-rate or bookings amount, or a share of revenue.</p> <p>▼ How it is measured</p> <p>Count each company in the frozen launch roster once if its last four reported quarters as of September 30, 2027 contain a company release, filing or company-hosted transcript with an aggregate AI-attributable revenue, ARR, run-rate or bookings amount, or a share of revenue. AI-native customer cohorts count when the company quantifies their revenue or ARR, but are separately identified: their whole bill is not necessarily AI workload. Growth rates, shares of growth, product adoption, customer counts, individual contract values and undifferentiated platform-wide consumption do not qualify. All 32 use this same rule. The target is 17; the source register distinguishes qualifying, excluded and unverified records. Unknown records remain unverified, not zero. Publish the underlying source and metric for every company counted at grading.</p>	September 30, 2027	Open
<p>B4 · A publicly evaluated model reaches an 80% task horizon of at least eight human-expert hours.</p> <p>▼ How it is measured</p> <p>METR's 80%-success horizon, central estimate, threshold 480 minutes. The horizon is the task duration, in human-expert time, at which METR's fitted curve gives the model an 80% chance of success; it is neither the longest task the model completed nor elapsed agent runtime. The grade reads the central estimate as METR publishes it at the deadline; a later change in METR's method does not reopen it. METR said on September 8, 2026 that its Time Horizon page is no longer actively updated; if no newer central estimate exists at the deadline, the grade reads the last published one and says so.</p>	June 30, 2027	Open
<p>B5 · OpenRouter's daily tokens triple.</p> <p>▼ How it is measured</p>	September 8, 2027	Open

BET	DEADLINE	STATUS
Complete-week average daily tokens against the registered baseline of 16.495 trillion a day (week marked September 6, 2026); threshold 49.49 trillion. Token volume is a usage indicator; it is not agent hours or spending.		
<p>B11 · The four bellwethers' combined revenue growth stays at or above 30% year over year in each of the next four reported quarters.</p> <p>▼ How it is measured</p> <p>Summed latest-quarter revenue of Cloudflare, MongoDB, Snowflake and Datadog against the same four fiscal quarters a year earlier, from SEC filings, for the four quarters reported after October 1, 2026: for Cloudflare and Datadog the quarters ending September 30 and December 31, 2026 and March 31 and June 30, 2027; for Snowflake and MongoDB the quarters ending October 31, 2026 and January 31, April 30 and July 31, 2027. Baseline 34.5% on the quarters ended June 30, 2026 (Cloudflare, Datadog) and July 31, 2026 (Snowflake, MongoDB). One quarter below 30% fails the bet.</p>	September 30, 2027	Open
Beyond 12 months · grading dates from January 2028 to March 2032		
BET	DEADLINE	STATUS
<p>B7 · AI agents performing user-directed work exceed human-originated requests in the disclosed traffic of at least two major platforms.</p> <p>▼ How it is measured</p> <p>A platform's own disclosure of its traffic by origin, with the same denominator in both readings (all requests the platform serves): requests from AI agents acting on a person's instruction, as the platform classifies them, against requests from people; ordinary bots, crawlers and scrapers are excluded from the agent share. A major platform is one of Cloudflare, GitHub, OpenRouter, Vercel and Stripe, or any platform disclosing more than one trillion requests a month; two of them must disclose the agent share above the human share. A research classification of a platform's volume, including ours, does not qualify.</p>	December 31, 2031	Open
<p>B8 · The 32-name backend reports more than \$250 billion of annual recognized revenue.</p> <p>▼ How it is measured</p> <p>Sum recognized revenue for fiscal years ending in calendar 2031 for the frozen 32-name launch roster, from company filings; this is a fiscal-year sum, not calendarized revenue. Translate Megaport at the average exchange rate for its fiscal year and publish the rate. Count a business only once: if acquired by another constituent, include it through the consolidated acquirer and add only the target's pre-consolidation revenue falling in the acquirer's reporting period, where separately disclosed. A target acquired outside the roster contributes only separately disclosed revenue for the relevant fiscal period; never carry its last four quarters forward as if they were 2031 revenue. Undisclosed, missing or incomparable amounts leave the grade unresolved. Baseline \$70.92 billion is the published 2025 fiscal-aligned sum. Threshold and March 31, 2032 grading deadline unchanged.</p>	March 31, 2032	Open
<p>B12 · Cloudflare, MongoDB, Snowflake and Datadog report more than \$40 billion of combined revenue for the 2029-aligned reporting periods, against the registered \$32.7 billion consensus snapshot.</p> <p>▼ How it is measured</p> <p>Reported revenue for calendar 2029 at Cloudflare and Datadog and fiscal years ending January 31, 2030 at Snowflake and MongoDB. Threshold: \$40 billion combined. The registered consensus comparator is \$32.7 billion from the September 8, 2026 snapshot;</p>	March 31, 2030	Open

BET	DEADLINE	STATUS
current consensus is displayed separately and does not reset it. Initial and final grade March 31, 2030. Missing filings leave the grade unresolved.		
<p>B14 · The backend is the best-performing layer of the stack over the three years from October 1, 2026.</p> <p>▼ How it is measured</p> <p>The same lines read at the October 1, 2029 close, from the September 30, 2026 close in the published index file, the start labelled Since October 1, 2026; the backend must show the largest gain. Launch roster rule: membership is frozen in launch-rosters.json. No new IPO or discretionary addition enters. Equal-weight daily returns use the same adjusted-close method as the published index. A delisting uses the last available adjusted close, then that position stays in cash at a zero return until the next trading day, when surviving members are equally reweighted; no replacement enters. If reliable delisting proceeds are unavailable, publish the missing treatment and leave the grade unresolved. Splits and dividends follow the adjusted-close series; acquisitions never add a new name. The launch-cohort grade must be published separately if the live tracker roster later changes. Graded separately as a market call; it overlaps with B1 and does not count as independent confirmation of the operating thesis.</p>	October 1, 2029	Open
<p>B15 · By the end of 2027, AI agents perform more than 100 million active hours a day worldwide.</p> <p>▼ How it is measured</p> <p>Observe the last complete week of December 2027, fixed even if evidence arrives later. Threshold: 100 million active agent hours a day worldwide, equivalent to 36.5 billion annualized hours. Read both the platform level and OpenRouter pulse under the dated Appendix A4 method. Publish both central estimates and assumption ranges. A robust hit requires both lower bounds above the threshold; a miss requires both upper bounds below it. Overlap, disagreement or missing evidence is unresolved. These are sensitivity ranges, not statistical confidence intervals. No single leg overrules the other at final grading. Initial grade January 31, 2028; one restatement of that same week by June 30, 2028 is final, including an unresolved grade. Independently review and register the method and range construction by September 30, 2027. This is active duration, not equivalent human output or jobs displaced.</p>	January 31, 2028	Open
<p>Dated October 8, 2026. The launch slate, October 8, 2026: eleven bets with stated measures and grading dates. H1 is the headline prediction. A test of paid retention of agent-built software waits on a matched 90-day baseline; no numerical target is set until that evidence exists. Beyond 2031, three calls stay on the record: A usage-oriented infrastructure business exceeds \$100 billion of annual recognized revenue by December 31, 2036; The public backend universe reaches \$1 trillion of annual recognized revenue by December 31, 2038; An external index provider launches an agent-infrastructure index by December 31, 2040. The registered predictions are in Appendix D2.</p> <p>Each prediction has a measure and a deadline. We grade it against the named source and retain the original terms. The same record appears in summary Figure 0.4 and Part Four’s Figure 4.8.</p>		

Research commentary from Skycatcher, LLC, an exempt reporting adviser, not investment advice. The baskets are hypothetical, hindsight-selected and not investable; the scenario figures are our estimates, not company guidance. Sources: SEC EDGAR, Financial Modeling Prep, METR, the GitHub Search API and OpenRouter. Nasdaq-100 is a registered trademark of Nasdaq, Inc., which does not sponsor or endorse this site. Full disclosures in [Appendix D3](#).