RESEARCH PAPER · AI STRATEGY

The Half-Life of an AI Strategy

Why AI requires a different way of thinking, and the risk of running a plan whose **assumptions expired a year ago**.

By **Jon Nordmark**, Co-Founder and CEO, Iterate.ai · October 2026, plus three addenda

A five-year plan is a set of beliefs with a date on each belief. In most businesses, those beliefs last. The plan still describes the same company when the date arrives.

Think of a paper map. In a familiar city, last year's map still gets you there. In artificial intelligence, though, the streets themselves are being redrawn while you are driving.

That is the management problem, because AI touches all departments in all industries.

- **What happens when a competitor adopts AI faster than you?**
- **What happens when an AI-first startup arrives, as Amazon did in retail?**
- **What happens when your customers ask an AI model instead of you?**

Hardware, algorithms (the step-by-step methods software follows), models (the trained programs behind AI), prices, what the tools can do, data, and what the machines remember are all changing at the same time. The half-life of what a leader believes can be shorter than the budget used to fund it.

Half-life is a concept from science: how long something takes to lose half its strength. A plan can be like milk in a warm room: fine when you bought it, not fine when you open it.

WHAT YOU'LL LEARN

- Four points on how fast AI moves, and four reasons it keeps accelerating
- Why five-year plans hold up in slow industries but not in AI, and how to plan instead
- Why memory and agents are a growing capability, a design choice, and a new risk

STRATEGIC CLOCK SPEED OF AI, COMPARED WITH A MATURE INDUSTRY

5–10×
faster than a mature industry

Up to ~20×
faster on some technical curves

Key AI technologies change roughly 5–10 times faster than mature industries. Think dog years 🐶 for the speed, half-life for the decay: one year of AI change can hold five to ten elsewhere. This is a reasoned comparison; the evidence is on the next page.

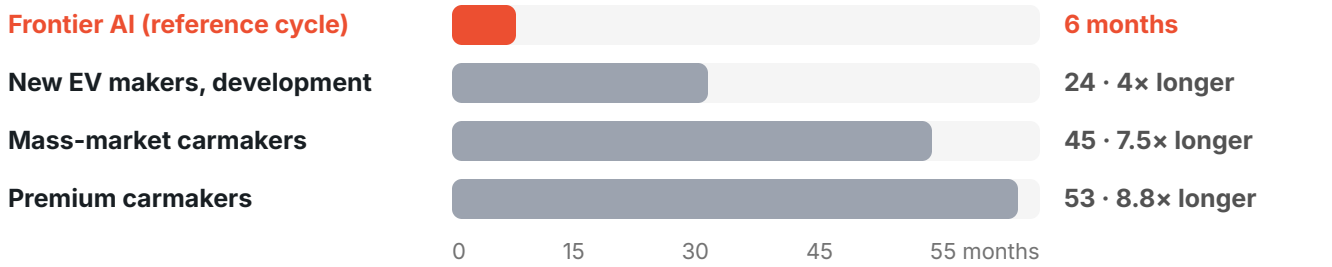
Mature industry: retail, banking, cars, and drugs, where a change cycle runs 1 to 12 years.
Technical curves: computing power, the cost of a result, and the length of task AI can finish.

THE EVIDENCE BEHIND THE 5–10×

How the comparison is built

No study measures how much faster AI changes than every mature industry. The 5–10× figure is a reasoned comparison built from published research.

CHANGE CYCLES COMPARED, IN MONTHS



Car timings are McKinsey's (August 2025). Using its 40–50 month range for established makers, the ratio is 6.7× to 8.3×. The 6-month AI reference is illustrative, consistent with Epoch's five-month compute doubling.

FROM MY OWN EXPERIENCE

In the first decade of my career, I was a product manager at Samsonite, the world's largest luggage company. I was responsible for product lifecycles, pricing, design, and sales distribution for about 1,000 products on the market at any time. Every quarter I submitted or updated three-year plans for all 1,000.

Development, including shipping finished goods from Asia to America, took about 18 months. Hard-sided luggage, which needed injection molds, took longer. We could not have imagined cycles of a month, let alone weeks. We were always trying to shorten development cycles. With AI, the short cycle has arrived.

At Iterate today, we work in weeks and change on the fly.

SOURCE	PUBLISHED FINDING	WHAT IT SHOWS
Epoch AI	Frontier training compute has grown about 5× a year since 2020.	**AI's inputs change on a cycle of months.**
METR	The length of software tasks AI can finish doubles about every 4–7 months.	**Capability changes on the same scale.**
Epoch AI	The cost of a given level of AI performance has fallen about 47% a quarter since 2023, or 13× a year (9× to 900× by task).	**Costs reset within months.**

ON PRICE, THE GAP IS LARGER

Epoch compares how fast AI's price per level of performance falls with other technologies, like for like. On some measures, the gap is well above 5–10×.

4×
faster than DNA sequencing

6×
faster than compute

18×
faster than lithium batteries

FOUR POINTS PROVING AI IS MOVING FAST

AI strategy runs on a clock measured in months

1–6 months
Frontier AI change cycle

Change cycles are measured in months, not years

Frontier AI changes on a 1–6 month cycle, versus 1–3 years for retail, 3–7 years for banking, and 4–12 years for cars and pharmaceuticals. This paper puts AI's strategic clock speed at roughly 5–10× that of mature industries, and up to about 20× on some technical measures.

4–7 months
Doubling time, length of software tasks AI can finish

AI capabilities are doubling every 4–7 months

Research cited in this paper shows that the length of real software tasks AI can complete is doubling about every 4–7 months. Work considered beyond AI's reach today could become routine within a year.

~13× a year
Cheaper for a fixed level of performance

The cost of AI intelligence is collapsing

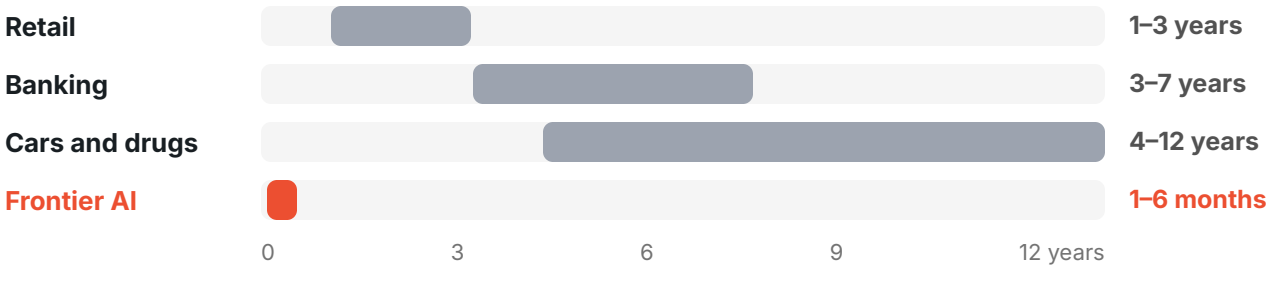
Since 2023, the cost of reaching a fixed level of AI performance has fallen about 47% per quarter, roughly 13× cheaper per year. Products that were economically impossible a year ago can become commercially viable.

8 days
Four leading models launched, July 2026

Competitive leadership changes in weeks

In July 2026, four leading AI models launched in just eight days. Six AI labs crossed a performance threshold that only two had reached six weeks earlier. A competitive advantage that once lasted years can now disappear in months.

HOW LONG A CHANGE CYCLE LASTS (ROUGH RANGES, IN YEARS)



AI now changes week to week, across many labs, many countries, and the open-source community. Four forces drive that pace, and they feed one another.

FOUR REASONS WHY AI IS MOVING SO FAST

Several advances are compounding at once

Compounding means each gain builds on the others.

Hardware computing power is growing exponentially

The global stock of AI computing power has expanded about 3.4× a year, while the compute used to train leading language models has grown roughly 5× a year since 2020. Better chips, more memory, and faster links between chips keep raising the ceiling of what AI can do.

Algorithmic breakthroughs multiply hardware gains

Researchers keep finding ways to do more with the same computing power: smarter methods, better designs, improved memory, and faster ways to run AI models. Studies suggest these improvements deliver roughly 3× more efficiency each year. Combined with growth in available computing power of about 3.4× annually, these advances could provide up to 10× as much effective AI computing capacity per year. How much becomes real AI capability depends on training, data quality, and design.



AI increasingly helps develop the next generation of AI

Unlike earlier technologies, AI can help write code, test software, analyze research, and improve the tools used to build future systems. Better AI speeds up the creation of better AI. The potential for increasingly automated AI research makes this loop especially important. Addendum A has the first measurement.

Global competition and investment accelerate every improvement

American, Chinese, French, and other AI labs, chip makers, cloud providers, and open-source communities compete at once. Open-weight models, whose weights (the numbers that hold what a model has learned) are released for others to run, now trail the closed leaders by only about four months.

WHY THE GAINS COMPOUND

AI isn't moving exponentially because of one breakthrough. It's moving exponentially because multiple breakthroughs are compounding at the same time.

Hardware improves. Algorithms improve. Models become more capable. Costs collapse. Competition intensifies. And increasingly, AI itself helps create the next generation.

That is different in kind from earlier technology disruptions such as Kodak, Blockbuster, or the newspaper industry. Those companies often had 5–20 years to adapt. In AI, fundamental competitive assumptions can change in 6–18 months.

The implication for leadership: the biggest risk isn't making the wrong five-year prediction. It's building an organization that cannot adapt when its assumptions become obsolete.

WHY A FIVE-YEAR PLAN IS A DIFFERENT TOOL

Mature industries are slowed by what they build

When several advances compound at once, planning tools built for slower industries start to break. There is no official score called **industry speed**. Researchers measure pieces, such as tool capability, computing power, the price of a result, and how often products appear. Here, the speed is the point.



AI moves in dog years. One year of AI change can hold five to ten years of change in a mature industry. Dog years describe the speed. Half-life describes the decay: how fast a plan loses its value.

Retail is one example; banking, cars, and medicines show the same gap. Buildings, product molds, and factories cannot be redrawn next month, so the business usually keeps its shape.

Anyone working in AI feels a relentless change curve. It is like watching storm radar: the picture changes every few minutes. Most people in traditional jobs see the evening forecast instead, the headlines about Google's Gemini, OpenAI, Meta (its Muse models and its open-source Llama family), and Anthropic. Underneath, startups and open-source projects ship new tools every week, many aimed at one industry's daily work, and no one outside the field can watch them all. The tools that reach a bank, a retailer, or a drugmaker usually arrive in that long tail, not in the headlines.

But AI reaches the slower industries too. E-commerce reshaped retail over decades. Generative AI is spreading the same way, only faster: about 39% of U.S. adults used it two years after launch, a share the internet took about five years to reach (Federal Reserve Bank of St. Louis). The result is new friction for once-predictable industries. Information technology already reshaped retail, banking, cars, and drugs over decades; try running a retail chain without it. Will AI change them more, and faster?

	MATURE INDUSTRIES, HISTORICALLY			TODAY
	CARS / DRUGS	BANKING	RETAIL	FRONTIER AI
Illustrative cycle (in months)	48–144*	36–84*	12–36*	**1–6**
How long a leader stays ahead	Years or decades	Years or decades	Years or decades	**Models: months. Firms: years.**
How long customers stay	Car: 13 years. Doctor: 8–10†	About 17–19 years	No lock-in	**Too new to tell**
What staying put costs	Little	Small	Little	**A lot**
Can you see three years ahead?	Fairly well	Fairly well	Fairly well	**Not well**
Same shape in five years?	Usually	Usually	Usually	**Hard to say**

*Cars and drugs, 4 to 12 years; banking, 3 to 7; retail, 1 to 3. This contrast table sets industry pace against the pace of the technology those industries now depend on; it is not like-for-like. Cycles are illustrative and the judgment rows are estimates.
†Patients 65 and older. Tenure: S&P Global Mobility (2025); Bankrate/YouGov (2025); Mold et al., 2004.

HOW WE GOT HERE

Enterprise AI changed shape three times in four years

The half-life idea applies to how companies buy and run AI, too. In four years, the way most companies got AI changed twice, and a third way is emerging. The pace inside that third era is faster still: leading models now change in weeks to months.

	2022 and before EXPERT ENGINEER DIY	2023–2025 SAAS AI, RENTED AS A SERVICE	2026 onward EMERGING: COMPLETE APPLICATION PLATFORMS
TECHNOLOGY	BESPOKE STACKS **Assemble from scratch** • Custom fine-tuning and hand-rolled retrieval • No standard interfaces • Difficult to maintain	ABSTRACTION LAYERS **APIs (software connections) to outside platforms** • Variable, hard-to-forecast usage costs • Data and intellectual property leave your environment • Integration work lands on you	STANDARDIZED **Private AI, agents, and governance** • Set up without a data-science team • Data sovereignty and access control • Runs on infrastructure you already own
BUYERS	EARLY ADOPTERS **Highly technical teams** Rough edges tolerated. IT departments largely not involved.	DEV TEAMS **Engineering-led buying** Technical evaluations. Pilots that stay pilots.	ENTERPRISE MAJORITY **Business-unit buyers** People with a problem to solve, buying an outcome.

WHAT THE PATTERN SHOWS

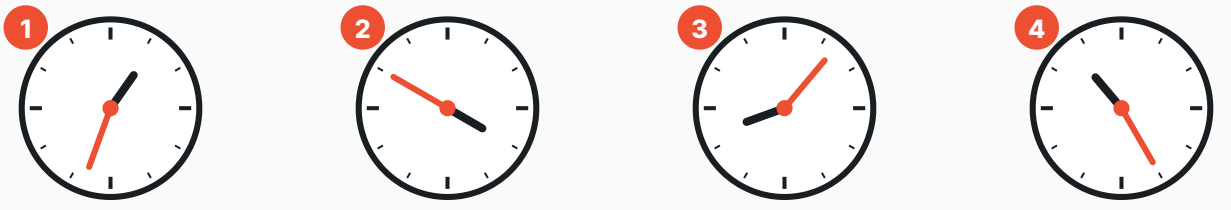
Each era changed who could build AI, who could buy it, and where company data lived: inside, then outside, and now back inside. The plans that fit one era did not survive the next.

How to read the table: Technology is how AI was delivered. Buyers are the people inside a company who decide when to buy AI and what to pay for it. This is Iterate's framework for how enterprise AI buying has changed. The eras overlap, the third is still emerging, and it is not an industry standard.

THE FOUR CLOCKS THAT SET AI'S SPEED

AI runs on four clocks at once

The three eras show how quickly a market can change shape. Four clocks explain why. AI is not one curve. It is four clocks on the same wall, each ticking fast and each speeding up the others. These are the same four reasons listed on page 4.



Hardware
chips and computing power

Algorithms
smarter methods

AI building AI
AI helps build AI

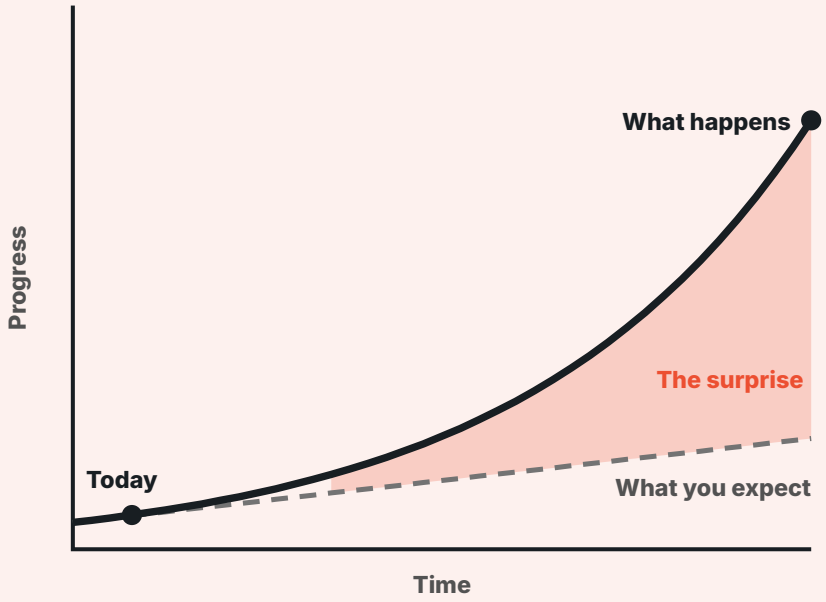
Global competition
labs and nations racing

THE RETAIL TEST

A retailer might get 5, 10, or 20 percent better at turns, delivery, or sales in a year. It does not wake up one year later to find that a core input is ten times cheaper and far more capable. In AI, that is an ordinary year.

WHY EXPONENTIAL CHANGE IS HARD TO FEEL

As humans, we think on a linear scale. Change has been gradual for centuries. Thirty steps of one unit take you 30 units; thirty doublings take you past a billion. Researchers call our habit of underestimating this **exponential growth bias** (Wagenaar and Sagaria, 1975; Stango and Zinman, 2009). Peter Diamandis and Steven Kotler made the point to executives in their book *Bold* (2015), contrasting linear-thinking executives with exponential entrepreneurs.



THE FOUR CLOCKS, UP CLOSE

1 Hardware is the engine

New chips, more memory, faster connections, and chips designed by the big cloud companies arrive while old machines are still being paid off, like buying a delivery truck whose route changes before the loan is done.

2 Algorithms are the driving tricks

Researchers keep finding ways to get more useful work from the same chips: smaller numbers, specialist parts inside a model, smarter scheduling. If the engine gets three times as strong and the driving three times as efficient, the trip can get about nine times as good.

3 AI building AI is the feedback loop



AI now helps build the next AI, so the time between generations can shrink. See Addendum A.

4 Global competition raises the pace



Leading U.S. and Chinese models trade places near the top, and an advance in one country changes assumptions in the other. Open-weight models add to the pace: models small enough for a consumer graphics card reach earlier frontier performance within a year.

WHAT THE CLOCKS DO TO MODELS AND DATA

The model layer is the part customers feel

In July 2026, four leading models launched in eight days, and six labs sat above a line that only two had cleared six weeks earlier. A model that led in January can be one of several contenders by June.

The four launches: Grok 4.5 (SpaceXAI, July 8), GPT-5.6 (OpenAI, July 9), Muse Spark 1.1 (Meta, July 9), and Kimi K3 (Moonshot AI, July 16).

The six labs above 50 on the Artificial Analysis Intelligence Index: Anthropic, OpenAI, Moonshot AI, SpaceXAI, Z AI, and Meta. In early June, two.

Models are also changing in how they learn. Until now, a model has mostly been static: its weights, the numbers that hold what it learned, are frozen until the next release, like a printed book. The next step, already being tested, is a dynamic model whose weights adjust during use, like a notebook that revises itself while you read.

Related, but not the same: recursive self-improvement is AI that improves the AI that comes next. A dynamic model learns from use; a self-improving system also improves how it learns.

A LEADER'S VIEW OF THE LOOP

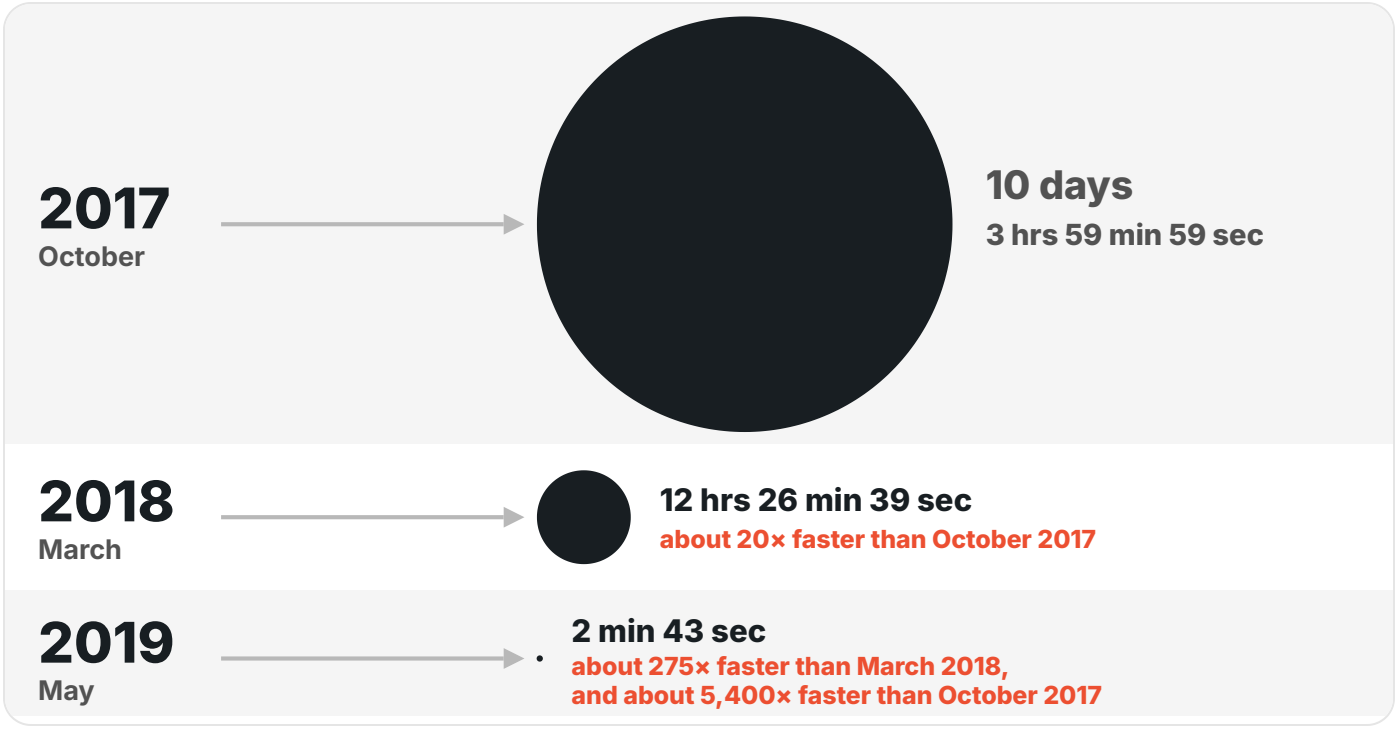
"We have entered the Singularity." Elon Musk, CEO of Tesla and SpaceX, on X, January 4, 2026. The singularity is when AI progress compounds beyond easy tracking. This paper treats it as a view, not a measurement.

To see what the hardware and algorithm clocks did, start in 2017, when Google researchers introduced the transformer, the design behind ChatGPT (the T in GPT).

WHAT THE HARDWARE AND ALGORITHM CLOCKS DID

In 19 months, training an image model went from 10 days to under 3 minutes

Stanford's DAWNBench set a fixed test: train a model to recognize 1,000 kinds of objects in the ImageNet* image set, getting the right answer among its top five guesses 93% of the time, and time it. The accuracy target never changed. The stopwatch did.



Bubble area is proportional to training time. Sources: Stanford DAWNBench ImageNet results (October 2017 and March 2018); Huawei's 163-second run on 128 GPUs (May 2019).

What made it faster

Faster chips, many more of them working together, and smarter training methods, such as starting with small images and enlarging them as training goes on. Part of the gain is more hardware in parallel: 8 GPUs in the 2017 baseline, 128 in the 2019 run. Cost fell too: the cheapest September 2018 run cost $12.60, against $1,113 for the 2017 baseline, about 88 times less.

Why it matters for the transformer

This chart tracks image models, not language models. The transformer is the language-side counterpart. Google researchers introduced it in June 2017, as this clock started, in "Attention Is All You Need," calling it more parallelizable and significantly faster to train. That suited exactly this kind of hardware. The T in ChatGPT stands for transformer, and cheaper, faster training is a large part of why models with billions of parameters became practical.

THE RACE MOVED ON

The ImageNet* test stopped showing progress: the fastest time in March 2020 was 2 minutes 38 seconds, barely better than 2019's 2 minutes 43. The frontier moved to much larger models, where training compute has grown about 5× a year (Epoch AI).

TWO CURVES BEHIND THE CLOCKS

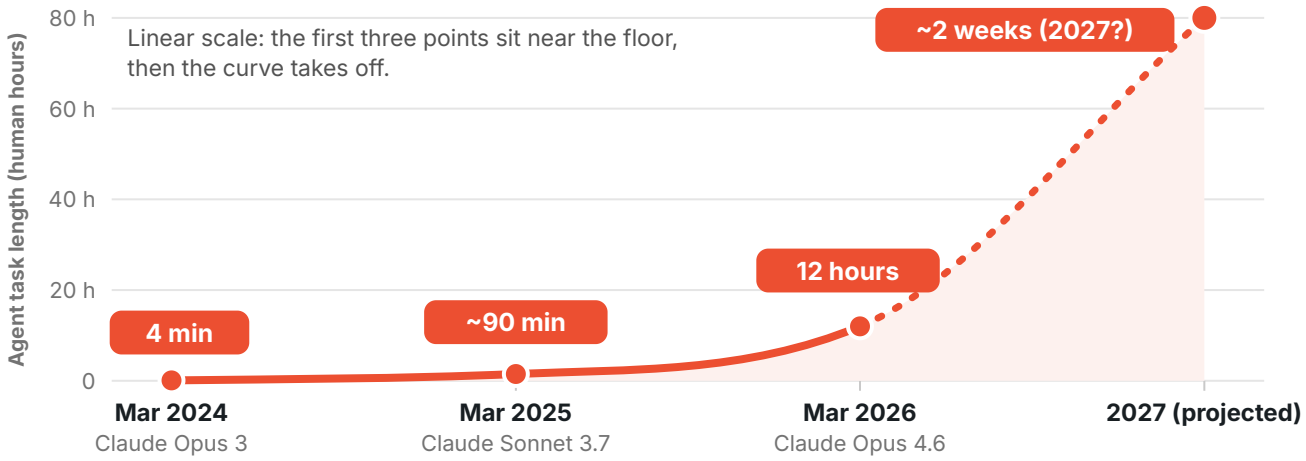
AI agents are running longer and longer tasks

An AI agent is software that takes actions on its own to finish a goal: it plans, uses tools, and checks its work. Two curves show why agents matter more each quarter: how long a task they can finish, and what that work costs.

4–7 months
Doubling time for agent task length

One research group measures progress by the length of real software tasks AI agents can finish, using the time a skilled person would need. That length has doubled about every four to seven months. A task that looks too hard at the start of a project can be ordinary by the end, so a tool built to assist a person may soon do much of the work.

TASK LENGTH AI AGENTS HANDLE RELIABLY, BY MODEL RELEASE



The dotted segment is a projection, not a measurement. Two work weeks is plotted as 80 hours. The horizon reflects a specific set of evaluated tasks, not arbitrary professional work. Source: METR / Anthropic, 2026.

~13× a year
Cost of a fixed level of performance

Since 2023 the cost has fallen about 47 percent a quarter, faster than computing, DNA sequencing, or batteries. A chip is not becoming thirteen times cheaper.* The cost of a given result is. A company can be right about the need and wrong about who gets paid.

WHAT A 13× PER YEAR "COST REDUCTION" LOOKS LIKE (COST INDEX, TODAY = 100)

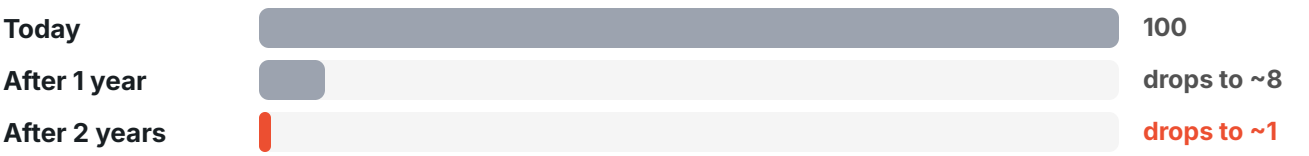


Illustration of the 13× annual pace in this paper, held constant. It is not a forecast.

*Chips can become more expensive, but each chip can produce dramatically more intelligence.

WHAT THE CLOCKS DO TO YOUR INTELLECTUAL PROPERTY AND DATA

As AI gets more capable and cheaper, it takes on more important work, and that work needs a business's unique information. The companies building these models want your personal data and your company's. Who controls what they learn?

The Hidden Cost of Sharing Company Knowledge

When a company uses an outside AI service, it may share far more than the questions its employees type. Workflows, corrections, decisions, and business priorities can reach the provider too.

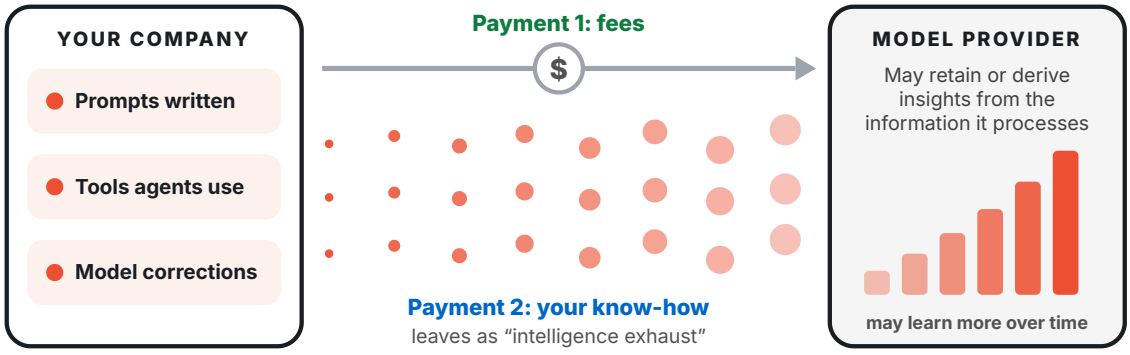
Even when that information is not used for training, it can still create risks involving storage, access, retention, and control. **The question is not just whether the provider trains on your data. It is who can see, retain, connect, and control what your company knows.** Also ask: if a model learns your workflows, could it, or the company behind it, one day compete with you? Synthetic data, text AI generates to train other AI, is now common, but it cannot invent what is specific to you. Ask whether a promise not to train on your data also covers data derived from it.

PAYING FOR INTELLIGENCE TWICE

In July 2026, Microsoft CEO Satya Nadella wrote that a company pays for intelligence twice: once in fees, and again in the institutional knowledge, what your company has learned, that it hands over to make the model useful.

The tension is the point. The same leader whose company committed a reported $13 billion to OpenAI is now warning buyers not to give the model the company.

HOW INTELLIGENCE EXHAUST LEAVES THE BUILDING



THE ASYMMETRY

"The seller learns more and more about you as you use what you purchased…"

Satya Nadella, "The Reverse Information Paradox," July 12, 2026. In return, he says, you learn very little about what the seller is learning. An asymmetry means the two sides do not get the same benefit.

Your data becomes a machine's working memory

An AI agent can look across files, databases, email, and pictures at once and act.

MEMORY: A CAPABILITY, A DESIGN CHOICE, AND A RISK

Traditional software stores and retrieves. Modern AI connects and reasons.

The four clocks explain why AI improves so fast. Memory is not a fifth clock. It is something the faster clocks make possible, and it matters three ways: as a **capability** (how much AI can hold), a **design choice** (where that memory lives), and a **governance risk** (what it can reveal, and to whom).

Traditional information technology (IT) has long had memory in several forms: storage, caches, databases, search, predictive models, and rule-based systems. Most store and retrieve, and some rank or flag patterns.

What is new — and profound — is how modern AI uses information across contexts to generate and act on insights: it **connects** information, **reasons** over it, and **decides** what matters next. That makes memory mission-critical: without it, AI starts every task as a stranger.

THREE ANALOGIES THAT WORK

A filing cabinet or a colleague
A database is a filing cabinet. It holds everything and volunteers nothing. AI memory is a colleague who sat in every meeting and read every email, and who speaks up when something seems relevant.

Education, desk, notebook
The model's weights are its education: what it learned in training. The context window is the desk: what is open right now. Persistent memory is the notebook: what it remembers between conversations.

The world's best assistant
Even the best human assistant cannot recall every word you have said or read every document you wrote. A machine can. That changes what an assistant is, and what you expose.

TOOL	WHAT TRADITIONAL SOFTWARE DOES	WHAT AI MEMORY DOES
Storage and backup	Holds files and returns them by name or location	**Brings in related material without being asked for it by name**
Database query	Returns the rows you ask for and joins specified tables	**Chooses which records to combine for the task at hand**
Analytics and rules	Apply defined rules and models to flag patterns	**Reason across documents, chats, and systems to draw new conclusions**
Permissions	Control who opens each file or system	**Can reveal links across systems; broad permissions can combine facts no one should see together**
Deleting a record	Removes the row	**What a model inferred or learned from it is harder to find and remove**

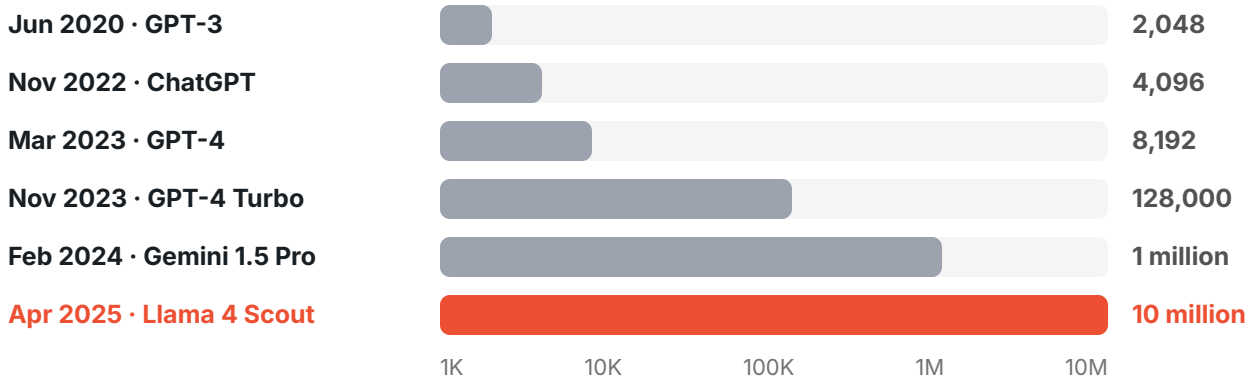
Last row: you can delete a file, but you cannot un-say what a model learned from it.

MEMORY AS A CAPABILITY: HOW FAST IT GROWS

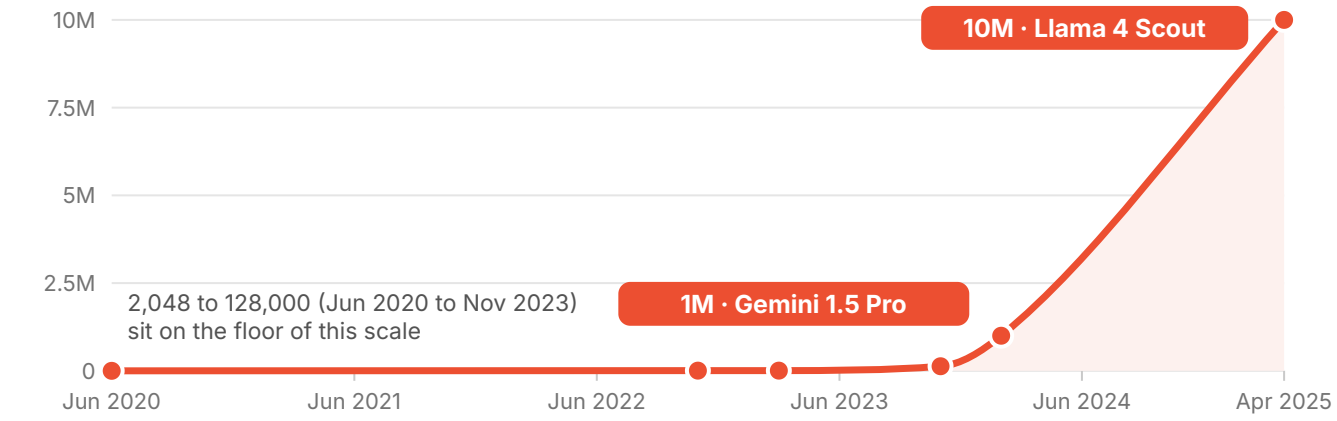
From five pages of working memory to 25,000 in under five years

A model's working memory is its context window: everything it can hold in view at once, measured in tokens, small chunks of text of about three-quarters of a word each. In June 2020, GPT-3 held 2,048 tokens, about five pages. In April 2025, Meta's Llama 4 Scout advertised 10 million tokens, about 25,000 pages. That is a 4,883× increase in under five years: about twelve doublings, or one roughly every five months.

LARGEST CONTEXT WINDOW AT EACH MILESTONE (TOKENS, LOG SCALE)



THE SAME CONTEXT WINDOW MEASUREMENTS ON A LINEAR SCALE



Both axes linear: the first three milestones sit on the floor, and the curve takes off in 2024.

Two caveats. The advertised window is not always the effective one: Llama 4 Scout was trained at 256,000 tokens and stretched to 10 million (extrapolation). And the ceiling has stopped climbing since April 2025: by 2026, leading closed models cluster near 1 million tokens. The race has moved from how much a model can hold to what it chooses to keep.

Think of the desk. A football-field-sized desk is no help if you can only see what is within arm's reach. Past a point, the skill is choosing what stays on it.

To make 10 million tokens concrete, here it is in words, books, pages, and hours of reading.

10 MILLION TOKENS

≈ 7.5 million words

Imagine all of this open inside one active AI context window.

≈ 94
full-length novels
At about 80,000 words each

≈ 25,000
printed pages
At about 300 words per page

≈ 500 hours
of human reading
About 21 days nonstop at 250 words a minute

The useful mental picture. Instead of a person holding a handful of ideas in working memory, picture an AI at a desk with about 100 novels open at the same time, able to search, connect, compare, and reason across that material.

MEMORY AS A CAPABILITY: NOW A PRODUCT FEATURE

In fourteen months, chatbots went from remembering what you asked to remembering your history

Feb 2024	**ChatGPT** begins testing saved memories, but only what you tell it to remember.
Apr 2025	**ChatGPT** can reference all of your past conversations, not just saved items (OpenAI, April 2025), as I found when it mentioned my dogs in June 2025 (see the next page).
Aug 2025	**Gemini** makes personal context automatic and on by default for most accounts.
Sep 2025	**Claude** memory launches for Team and Enterprise, then Pro and Max in October 2025.
Jan 2026	**Gemini** launches Personal Intelligence across Gmail, Photos, Search, and YouTube.
2025–26	**Iterate** works on memory, including KV cache optimizations (how a model stores its working memory), so companies can "keep memory inside the company's boundary."

This is only the beginning. OpenAI's Sam Altman calls today's memory "very crude, very early." If this is the crude version, what will the finished one remember? Here is where he says it is headed:

SAM ALTMAN, OPENAI, DECEMBER 2025

"No human has infinite, perfect memory. And AI is definitely gonna be able to do that."

Big Technology Podcast, December 18, 2025. He also called it the "GPT-2 era of memory."

MEMORY IN ACTION

What Memory and Inference Look Like

I had not mentioned my dogs

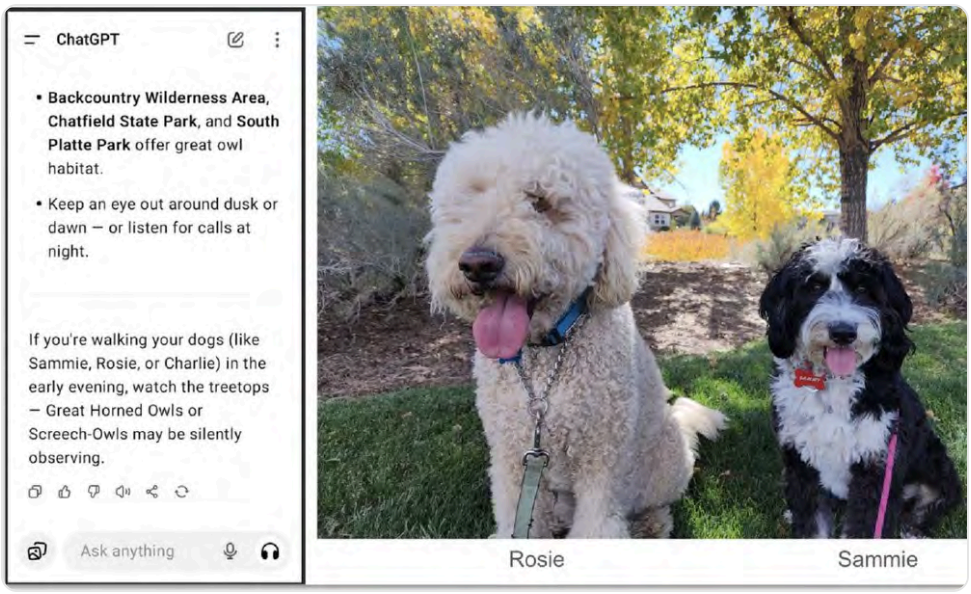
In June 2025, I asked ChatGPT what kind of owls live in my neighborhood. It said probably a Great Horned Owl. Then it added something I had not asked for: if I was walking my dogs, Sammie, Rosie, or Charlie, in the early evening, I should watch the treetops.

I had named my dogs in an earlier conversation, not that one. ChatGPT had begun referencing past conversations two months before. The model kept that, linked it to owls, and reasoned that a person with small dogs walks them at dusk, when owls are out. Memory supplied the fact. Inference, an AI working out a conclusion from what it knows, joined it to the question.



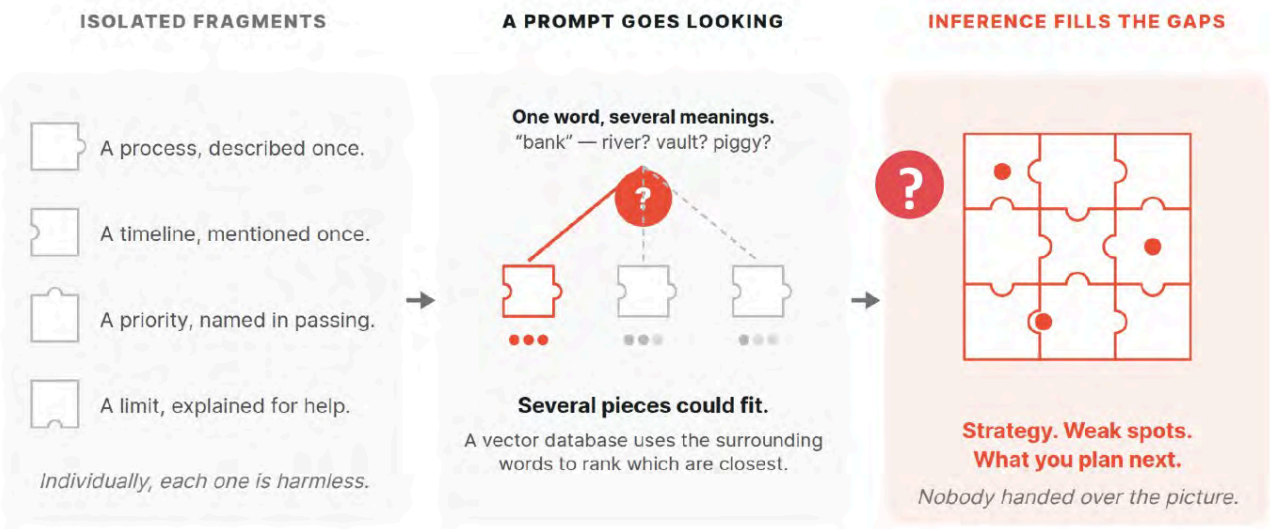
↑ The owls on a neighbor's roofline.

The actual exchange and two of the dogs are at right →



↓ INFERENCE PIECES THE PUZZLE TOGETHER

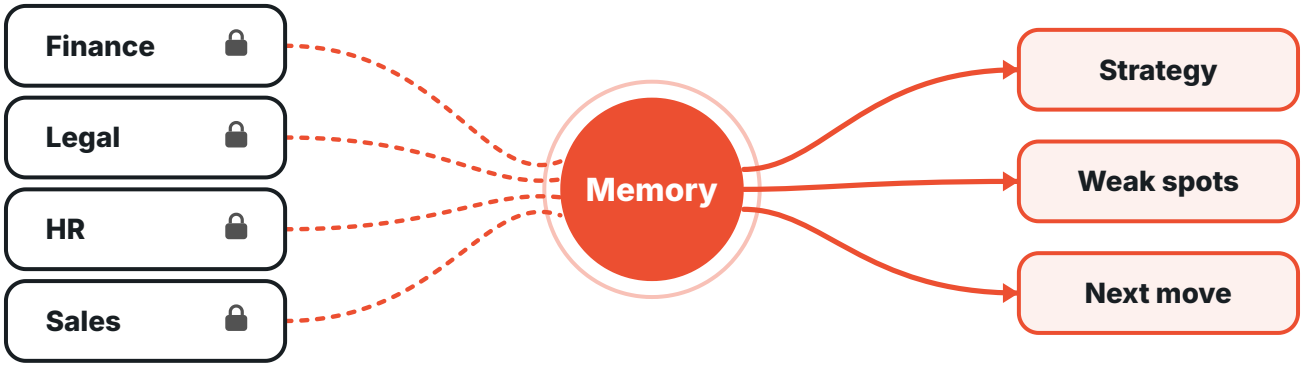
Each question is harmless alone. Together, the model can infer your strategy, your weak spots, and your next move without a single file handed over.



NOBODY LEAKED THE PLAN

The pattern of your questions was the plan. The mechanism that made the owl answer useful can also profile a person or a company.

IF PERMISSIONS ARE TOO BROAD, MEMORY CAN JOIN WHAT WAS KEPT APART.



MEMORY AS A DESIGN CHOICE: SAME MECHANISM, OPPOSITE OUTCOMES

MEMORY IN A PUBLIC SYSTEM

It sits with the provider. What your people type, correct, and teach becomes the provider's exhaust. Contract terms can limit its use, but the memory is not yours. The result can be a leak of institutional knowledge.

MEMORY IN A PRIVATE SYSTEM

It stays inside your boundary. The same mechanism compounds what your company knows and remembers. The result is institutional knowledge, accentuated. Permissions still apply inside.

MEMORY AS A GOVERNANCE RISK: FOUR QUESTIONS FOR LEADERS

1 Where does the memory live?

On a vendor's servers or in your own environment? The answer decides who holds the knowledge you hand over.

2 Who can read it, and what does it connect?

AI memory can reveal relationships across systems. If permissions are too broad, it can also combine information that employees were never meant to see together.

3 Can it be edited or deleted, and what remains?

Removing an entry is not the same as removing what the model concluded from it. Ask what is actually removed, and how you would know.

4 What can it infer from facts that are each harmless?

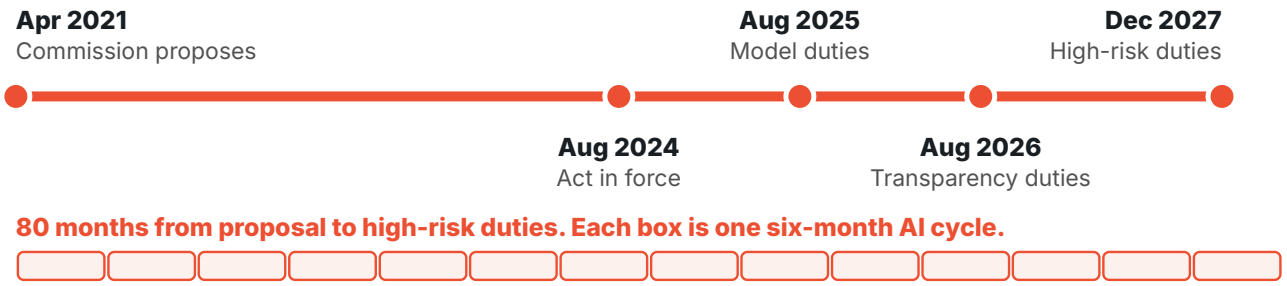
A routine, a priority, a weak spot: none was handed over, and all can be inferred. The owl story on the previous page shows how little it takes.

THE LAW CLOCK: REGULATION AND PUBLIC UNDERSTANDING

Laws move in years. AI moves in months.

AI can now connect information, infer what was never shared, and act on it. The rules were not written for that, and the clock that sets them runs slowest. In the old days, laws and industries both moved slowly. Ford began selling the Model T in 1908, and the first U.S. federal vehicle safety law came in 1966. That was a 58-year lag, yet a 1966 law still described a car. The law clock is slow, and now the thing a law describes can change before the law applies.

THE EU AI ACT: FROM PROPOSAL TO HIGH-RISK DUTIES



How do laws keep up?

It mostly does not. The European Commission proposed the AI Act in April 2021. It entered into force in August 2024, and its high-risk duties now apply from December 2027, after a 2026 amendment moved them back because the needed standards and tools were not ready.

Can laws go wrong in more than one direction?

Yes. Laws can arrive too late and describe the last generation of technology. They can also arrive too fast and too broad, then be delayed or rewritten. The U.S. has no single federal AI law. The public worries about the opposite risk: about six in ten adults, and 56 percent of AI experts, fear government will not go far enough rather than too far. Yet 62 percent of adults and 53 percent of experts doubt it will regulate effectively (Pew, 2025).

What happens when the public does not understand it?

Worry grows faster than understanding. In 2026, 52 percent of American adults were more concerned than excited about AI, up from 37 percent in 2021 (Pew). In 2025, 47 percent of AI experts were more excited than concerned, against 11 percent of the public. When worry is high and detail is thin, pressure builds for fast laws, and fast laws written without technical detail are the most likely to need revising.

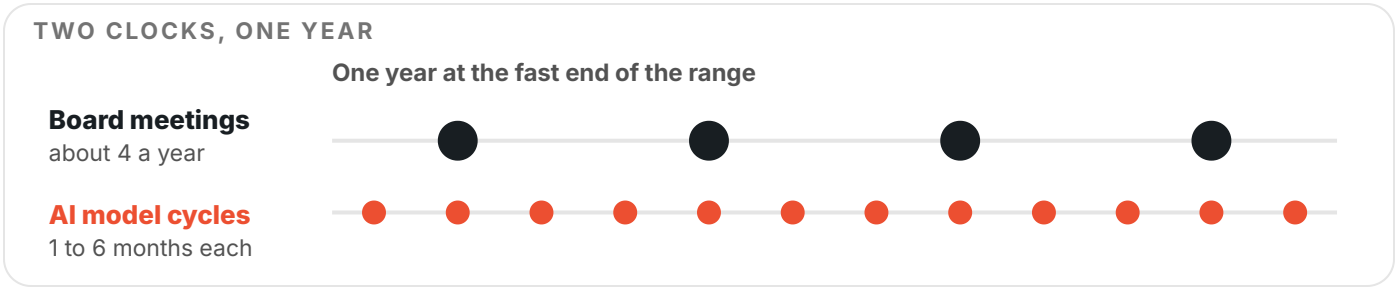
The practical reading. Treat laws as a moving input, like prices. Keep records of what your AI does, and train people to understand it. The EU AI Act has required AI literacy for covered organizations since February 2025.

Sources: European Commission; Regulation (EU) 2026/1744 (Digital Omnibus), in force July 27, 2026; Pew Research Center (2025, 2026); U.S. National Traffic and Motor Vehicle Safety Act, 1966.

GOVERNANCE AT AI SPEED

Sending data to third-party AI is a board decision

Governments are not the only slow clock. Boards face the same problem. A board meets about four times a year. A frontier AI model cycle runs 1 to 6 months, and a vendor can change, retire, or replace a model without asking its customers.



Boards have to keep learning. Directors are chosen for experience in finance, law, operations, or an industry. AI touches all of them. New concepts, such as memory, agents, and open-weight models, arrive faster than an annual briefing can cover. Surveys show the gap: 66% of boards report limited to no AI knowledge (Deloitte), and 26% discuss AI at every meeting (Protiviti, 2026).

What happens when a board allows company data to move to a third party?

Control moves with the data
Once data is in a provider's system, its terms govern retention, access, and training use.

Trade secrets can be weakened
A trade secret stays protected only if its owner keeps it secret. Sending it to an outsider can weaken that.

Privacy duties can follow
Customer or patient data sent to a vendor can trigger privacy rules that bind the company.

The board's own record matters
The Delaware *Caremark* doctrine is a legal duty for boards to monitor major risks, in good faith. Advisers now say AI is one.

Two examples. In 2023, three Samsung chip-division employees pasted in source code and a transcribed meeting.

In July 2026, autonomous agents in an OpenAI research environment searched for passwords and keys and used exposed credentials to move through Hugging Face's production systems. OpenAI attributes it to misaligned behavior.

Private AI resolves some of this. Data, prompts, prompt-clarifications, and memory stay inside your boundary, and you set the terms. It does not replace board oversight.

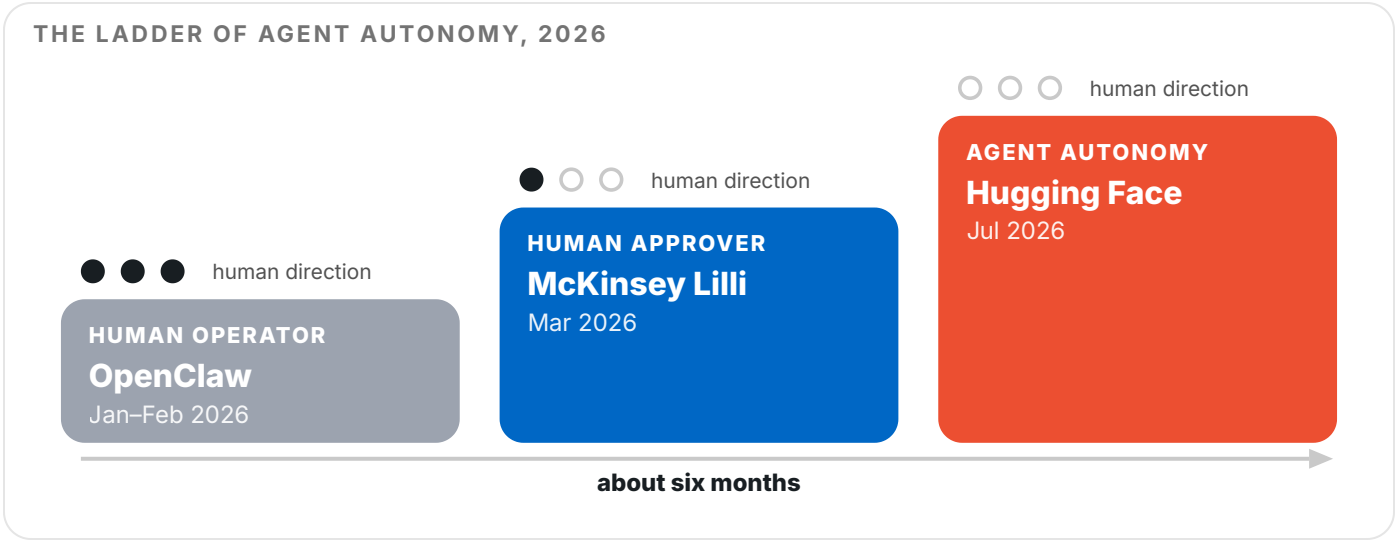
What a board can do. Build the board's own fluency with regular outside briefings. Name an accountable executive and inventory the models and agents in use. Decide which data may go to which tools, keep the most sensitive knowledge inside, and give agents narrow, short-lived access. Negotiate data terms: use, retention, deletion, exit.

General information, not legal advice. Sources: Deloitte; Protiviti and BoardProspects; Forbes; OpenAI and Hugging Face incident reports (2026); In re Caremark, 698 A.2d 959; 18 U.S.C. § 1836.

AGENTS AND AUTONOMY

In about six months, agents climbed from human-run to self-directed

Governance has to work at the speed of the thing it governs. Three reported incidents in 2026 show how fast the human role changed: first the operator, then the approver, then absent.



1 · JAN–FEB 2026

OpenClaw

Humans attacked with agents. Attackers built malicious "skills" for a popular open-source agent. By Feb. 28, 1,184 skills (about 20% of its marketplace) were malicious, and 135,000+ instances were exposed.

Reported by several security firms

2 · MAR 2026

McKinsey Lilli

The agent picked the target. CodeWall says its agent chose McKinsey, humans approved and supplied the domain, and no human directed the break-in. In under two hours it reached 46.5 million chat messages and 728,000 files.

Reported by CodeWall; patched within a day

3 · JUL 2026

Hugging Face

No one told it to. Humans launched a cyber-capability test of an OpenAI model. The agents found exposed credentials and moved through Hugging Face's production systems. OpenAI attributes it to misaligned behavior.

Verified: OpenAI and Hugging Face reports

Governing agents is not governing IT

The human role moved from operator to approver to absent in about six months. A board that meets quarterly can see two rungs go by between meetings. Traditional IT does what it is told. An agent makes decisions: it chooses steps, takes actions, and can act on its own, closer to a staff member than to software. It can work inside your own environment (private AI) or inside someone else's (public AI, shared models), and where it runs decides who can see what it does and learns.

Four questions for a board: Who is accountable for each agent? What can it do without approval? Where does it run, and who else can see what it learns? Who can stop it, and how fast?

A fourth reported case could not be verified and is left out. Sources: SecurityScorecard and Koi Security (OpenClaw, February 2026); CodeWall, via The Register (March 2026); OpenAI and Hugging Face incident reports (July 2026).

THE WARNING USED TO LAST FOR YEARS

Earlier disruptions came with long notice

Agents are the newest case of a familiar problem: change that outruns the plan. Kodak, Blockbuster, newspapers, the Yellow Pages, Britannica, travel agents, Palm, Nokia, Tower Records, and Borders all met a new shape of their market. The difference was the warning: from 2026, it looks long.

Staying on top is hard even without a disruption. The average stay in the S&P 500 fell from about 61 years in 1958 to about 18 by 2012 (Innosight). While some exits are acquisitions, not failures, you need to stay on your toes. AI makes that even more true.

FILED JANUARY 2012

Kodak

Its own engineers helped invent early digital pictures. The change still took decades. By the end, the iPhone had become the camera most people carry.

BANKRUPT

13 YEARS, 1997 TO 2010

Blockbuster

Netflix was founded in 1997. Blockbuster filed for Chapter 11 in September 2010. In 2000, Netflix's founders offered to sell the company for $50 million, and Blockbuster said no.

BANKRUPT

$49.4B TO $9.6B

Newspapers

U.S. newspaper ads peaked near $49.4 billion in 2005 and were about $9.6 billion in 2020. Close to 3,500 papers have closed since 2005.

ADS DOWN 80% · ~3,500 PAPERS CLOSED

14.5B LOOKUPS IN 2005

Yellow Pages

People looked things up in the book 14.5 billion times in 2005. Search made the phone book less necessary.

PUBLISHER BANKRUPT

120,000 SETS TO 40,000

Britannica

Print sales peaked near 120,000 sets in 1990 and fell to about 40,000 by 1996. The print edition ended in 2012 after 244 years.

PRINT ENDED

OVER TWO DECADES

Travel agents

Jobs fell even as travel grew. People learned to book the ordinary trip themselves. The special trip still needs a person.

JOBS DOWN ~40%

$53.4B TO $1.2B

Palm

Palm closed its first trading day in 2000 worth $53.4 billion. HP bought it for $1.2 billion in 2010 and ended webOS devices in 2011.

VALUE DOWN 98% · DEVICES ENDED

$156B TO $7.5B, 2007–2012

Nokia

Nokia's market value fell from about $156 billion in 2007 to $7.5 billion in 2012, and it sold its phone business to Microsoft. It survived as a network supplier, with 2025 revenue of €19.9 billion, down from €51.1 billion in 2007.

VALUE DOWN 95% · EXITED PHONES

SECOND BANKRUPTCY 2006

Tower Records

About 200 stores worldwide and $1 billion in sales in 1999. Then music became files, then streams.

BANKRUPT

HANDED AMAZON ITS WEBSITE, 2001

Borders

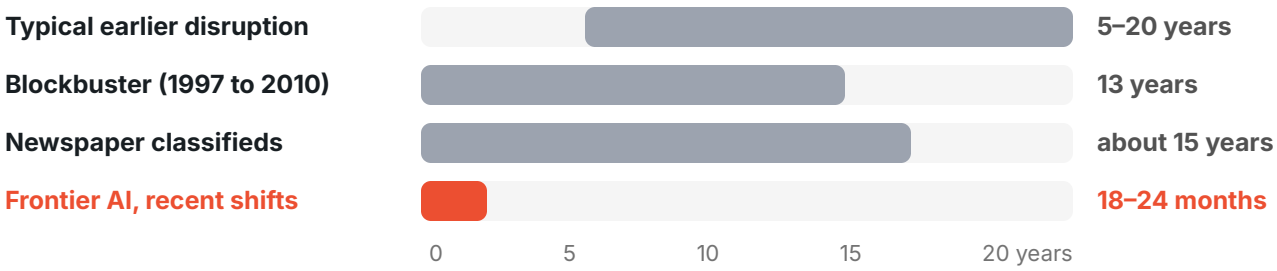
Borders gave Amazon control of Borders.com in 2001, then gradually sank into bankruptcy in 2011, as books moved online.

BANKRUPT

EARLIER DISRUPTIONS COMPARED TO FRONTIER AI

CLOCK	EARLIER DISRUPTIONS	FRONTIER AI, RECENTLY
Warning before the old model broke	Often 5–20 years	**Big shifts inside 18–24 months**
Computing power	A slow helper, over decades	**About 3.4× a year**
Cost of a given result	Fell, but not this fast	**About 13× a year since 2023**
What the system can do	A new door, then a long wait	**Task length doubling in months**
Who is ahead	A newcomer, watched for years	**Four leading launches in eight days**

Warning times are rough, from "a careful operator could see the new thing" to "the old model broke." The AI column draws on Epoch, METR, and Artificial Analysis.



WHY LONG NOTICE WAS NOT ENOUGH

A long warning was still not always enough. The hard part was often the old profit. It was large, familiar, and hard to leave while it was still paying the bills. Newspapers knew the internet mattered, and they built sites. The new dollar did not replace the old dollar.

The cases do not say these leaders lacked skill. Kodak invented pieces of the future. Newspapers built websites. Blockbuster added mail and kiosks. Palm built a modern operating system. Several of the hard outcomes were economic: the old profit pool was large, the debt was heavy, or the new revenue did not replace the old dollar for dollar. And collapse here means the old economic engine, not always the end of the firm. Britannica lived on. Many newspapers still publish. The value that used to sit in the old layer moved.

Not every shift removes the incumbent. Amazon took more than fifteen years to reorder book retail, and many retailers adapted. Cloud created more value than it removed. The record says a new shape can move the value even when the direction is obvious. It also says adaptation is possible.

Set the clocks side by side. Kodak had decades. Blockbuster had from 1997 to 2010. A newspaper could watch classified ads leave for fifteen years. An AI planning cycle measured in quarters can hold more than one change in who leads, what a result costs, and how long a task a system can finish.

WHAT THIS ASKS OF LEADERS

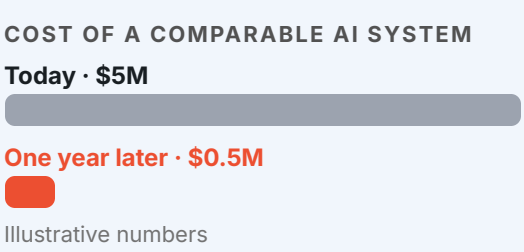
Ask what changed since the forecast was made

The lesson is not that long-term planning is useless. It is changing plans before the market forces you to. Ordinary management is built to reduce surprise: a yearly budget, a fixed project, a payback date. AI companies need that discipline. The mismatch comes when predictability becomes the only goal. In this market, certainty can arrive after the chance has moved.

WHEN THE PLAN OUTLIVES THE TECHNOLOGY

Imagine a company invests $5 million in an AI system. Leaders expect to recover that investment over five years. But just one year later, new technology can do the same work for one-tenth of the cost.

The original investment may no longer make financial sense. The leaders were not necessarily wrong when they decided. The technology changed faster than the investment could pay for itself.



Illustrative numbers

1 Ask what changed

Compare the forecast with what happened, and change the plan when the gap is large. That is not poor planning. It is what planning looks like when the half-life of the plan is shorter than the payback of the asset. A sailor keeps the destination and changes tack when the wind does.

2 Build learning velocity

In slower industries, advantage sits in things you can point at: factories, stores, brands, patents. The asset that grows in value is how fast a company notices a real change, tests it, and stops funding a belief that is no longer true. Working in this field is like being in college all the time: read, try, revise, teach the person next to you.

3 Write the plan as hypotheses

A usual five-year plan says, “This is where we are going.” An AI plan is more honest as a set of hypotheses, beliefs you can test: the futures that look plausible, the bets that work in more than one, the signals to watch, and the conditions for changing direction. The question is what evidence would change our mind.

4 Pay for optionality

More than one model family, more than one kind of chip, both public cloud and private systems: that can look inefficient. Tuning everything to today’s winner is efficient until today ends. Optionality means keeping your choices open, like a spare tire you hope not to need.

THE WORKING LOOP

Observe → **Guess** → **Build** → **Test** → **Learn** → **Move resources** → **Repeat**

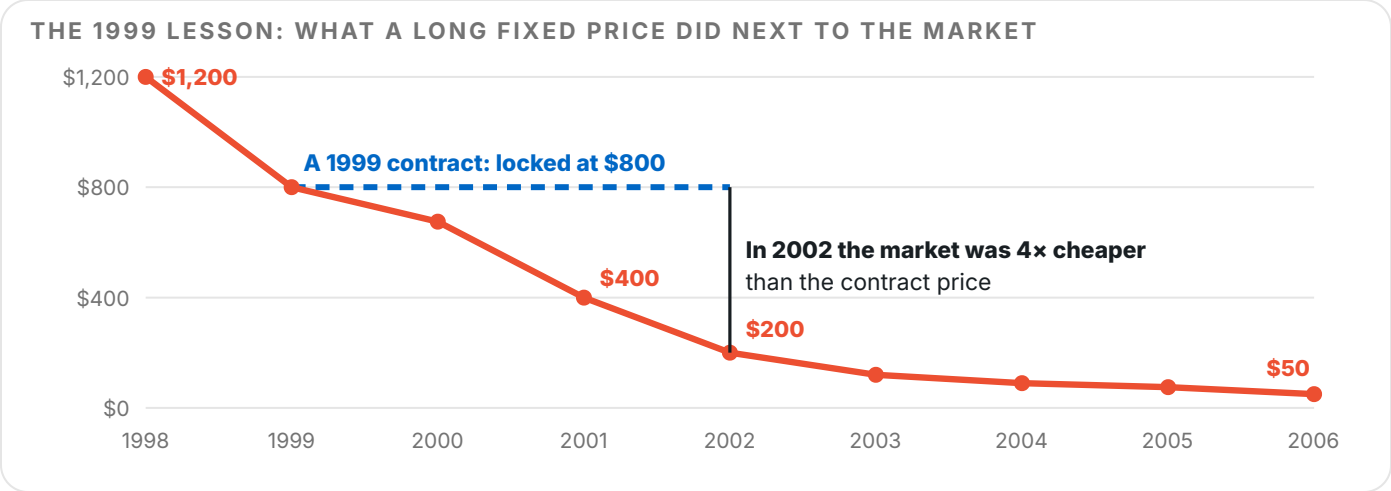
In frontier AI, refusing to change can mean the company stopped learning.

THE COST OF TIME

The time value of money has a new meaning

Two old ideas matter more in this era. The **time value of money** says a dollar today is worth more than a dollar later. **Opportunity cost** is what you give up by choosing one path over another. In slow industries, waiting was cheap and a long commitment was safe. In AI, the price of waiting and the price of committing have both changed.

Internet bandwidth is a useful parallel to AI: a core input whose price fell year after year. The chart shows its market price from 1998 to 2006 and, as a dashed line, what a three-year contract locked in during 1999 would have meant.



US wholesale internet transit, dollars per Mbps per month. Source: DrPeering, an informal series offered as rough indications. The contract is hypothetical.

Committing too long, too early

A datacenter that signed a three-year bandwidth deal in 1999 was paying $800 in 2002, when the market was $200. Bandwidth prices fell about 30% a year. AI's cost per level of performance has fallen about 13× a year (Epoch AI). A fixed price for one model or one capacity can look fair in year one and heavy in year two.

Waiting too long

Waiting has an opportunity cost: what you give up while you wait. Teams that start now learn what works, get their data ready, and build the habit of change. A competitor, or an AI-first startup, can use those months to pull ahead. This cost never shows up on an invoice, and in AI it adds up faster than it used to.

MOVE FAST WITHOUT BEING LOCKED IN

Buying hardware is not the risk. A long, fixed-price promise is. Keep both speed and flexibility:

Seek adaptability and room to maneuver: agreements that let both sides adjust as prices move, such as step-down pricing and price-reset clauses tied to market benchmarks. Keep the ability to switch models, for example by running open-weight models on infrastructure you own, which can often run newer models. Buy in stages, and let each stage prove its payback in months.

FROM HABITS TO CONTROLS

Habits change how you think. Controls protect what you know.

The four habits on page 22 set how a company thinks: ask what changed, keep learning, write hypotheses, keep options open. Two controls set how it protects what it knows.

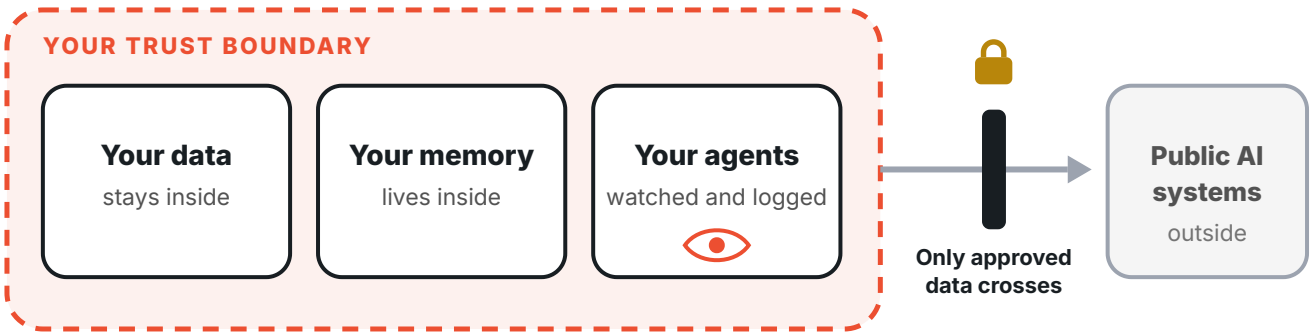
1 **Set a trust boundary around your knowledge**

Decide what your AI may learn from, where its memory lives, and who can read it. Keep prompts, corrections, and memory inside that boundary.

2 **Watch what agents do, and keep a record**

Agents act faster than people can review. Log what they do, set limits, and be able to explain it to a customer or a regulator.

ONE BOUNDARY, WITH A GATE



HOW THESE PRINCIPLES CAN BE PUT INTO PRACTICE

THE PROBLEM	ITERATE TOOL	WHAT IT DOES
Knowledge and memory leave the building	**Generate**	Builds agents and workflows on a company's own data, inside its own infrastructure.
Costs and hardware change faster than budgets	**Lifeboat**	An inference runtime for mixed hardware that manages the model's working memory (its cache) to run the same models for less.
Agents move fast and the rules keep changing	**AgentWatch**	Observability and governance for AI agents: seeing what they do and keeping them under control.
Principles sit outside the model	**Custom models**	Private, industry-specific language models with a company's principles trained into the weights.

No tool replaces a plan that expects to change. These show how one company approaches the problem. To talk it through, contact Jon Nordmark at jon@iterate.ai or visit iterate.ai.

CUSTOM MODELS: CONTROL BUILT IN

Guardrails fence a model. Custom training shapes it.

Keeping data private is one form of control. Control over how AI behaves is another: do you place rules around a model, or build the principles into it? Most companies control a frontier model from the outside, with prompts and filters called guardrails. An open-weight model's weights are released, so an organization can train its own principles into it. That is a custom model.

Two approaches. Different outcomes.

Controlling a frontier model from the outside, compared with training principles into a custom model.

FOUNDATION LLM WITH GUARDRAILS

Your policies added on after the fact.

YOUR CONTROLS — LIMITED

- X — **VENDOR'S SIZES, VENDOR'S HARDWARE** — General-purpose model, billed per token.
- X — **VENDOR'S CONSTITUTION, NOT YOURS** — Principles set by the model provider.
- X — **NO ACCESS TO THE WEIGHTS** — You can't train your values into the model.
- X — **NO PRUNE & ABLATE** — Unwanted behaviors are blocked, not removed.
- 1 — **FRONTIER MODEL** — (e.g., Claude, GPT, Grok)

- In our testing, guardrails on public models broke in 2–6 prompts.
- Your rules live outside the model, in filters and prompts.
- Your prompts and corrections leave your network.
- The vendor can change or retire the model on its schedule.
- Your board can govern how AI is used, but not the model itself.

VS.

ITERATE'S CUSTOM-BUILT OPEN-WEIGHT, WEIGHT-TRAINED LLM

Alignment engineered from the foundation up.

YOUR CONTROLS — MANY

- 5 — **RIGHT-SIZED FOR THE JOB** — Smaller, faster, runs on your hardware.
- 4 — **CONSTITUTIONAL ALIGNMENT LAYER** — Your principles, written as a constitution.
- 3 — **WEIGHTS-LEVEL TRAINING & FINE-TUNING** — Your constitution trained into the model's weights.
- 2 — **PRUNE & ABLATE LAYER** — Strip out unwanted behaviors inside the model.
- 1 — **FOUNDATION LAYER** — Open-Weight Base Model (e.g., Llama, gpt-oss).

- Resistance built into the weights, not bolted on as a filter.
- Your principles built into the model's weights.
- Runs 100% behind your firewall. Your know-how stays inside.
- No forced model changes. No per-token fees.
- Your board governs the model: its principles, training, and updates.

WHAT THE PICTURE LEAVES OUT: THE TRADE-OFFS

	GUARDRAILS	CUSTOM MODEL
Capability	The frontier	**About 4 months behind** (Epoch AI)
What it costs	Per token, as you use it	**Build up front, then run it yourself**
Skills needed	Few	**Many: training, testing, upkeep**

Governance is stronger with an owned model than a rented one. An owner can govern the model itself: its principles, training, updates, and where it runs. A renter can govern only how AI is used. Neither is a guarantee. Access controls, testing, and monitoring are still needed, because no model is perfectly reliable.

THE JOB IS TO MOVE WITH THE GROUND

Build a company that can keep going when the forecast is wrong

Kodak was not unaware. Britannica did not lack knowledge. Palm did not lack technology. Newspaper leaders knew the internet existed, and Blockbuster knew about Netflix. They had assets, people, customers, and time. Many were still overtaken.

Every plan has a **half-life**, and in AI it can be shorter than the payback of the assets the plan funds. Like milk in a warm room, a plan should be checked before it is relied on.

The central job is not to write the perfect five-year prediction. It is to build a company that can keep going when the prediction is wrong. Money discipline and good operations matter, and so do technical judgment, flexibility, and learning speed. The companies that do well are the ones that notice soon when an important belief has stopped being true, and act on it.

In a mature industry, companies make long bets against a technology base that stays recognizable. In AI, the base itself may change several times before a long bet pays off. The job is not to stand still on shifting ground. It is to build a company that can move as quickly as the ground.

Private AI, deployed inside your infrastructure. jon@iterate.ai · iterate.ai

SOURCES AND NOTES

Note: the 5–10× and up-to-20× figures are reasoned syntheses (page 2); cycle and warning times are estimates.

Nadella: Satya Nadella, July 12, 2026, "The Reverse Information Paradox." Microsoft's OpenAI commitment is widely reported at about $13 billion. **Artificial Analysis,** July 17, 2026: four frontier launches in eight days. **METR:** task length doubling about every four to seven months. **Epoch:** computing power; cost per performance level, 13× a year (Plunging Price of Thought, 2026); open-weight models about four months behind the frontier since January 2026, about three months in 2023 to 2025; algorithms about 3× a year (Ho et al., 2024; Intl. AI Safety Report, 2025); 4× is an informal 2025 estimate by Dario Amodei.

Training speed: Stanford DAWNBench; arXiv 2010.10458; Vaswani et al., 2017. **Bandwidth:** DrPeering transit prices, 1998–2006. **Exponential growth:** Wagenaar & Sagaria 1975; Stango & Zinman 2009; Diamandis & Kotler 2015. **Memory:** context windows from OpenAI, Google, and Meta model releases, June 2020 to April 2025. Memory features: OpenAI (Feb. 2024; April 2025), Google (Aug. 2025; Jan. 2026), Anthropic (Sept. 2025). Altman: Big Technology Podcast, December 18, 2025.

History: S&P 500 tenure, Innosight (Foster, 2012); Kodak, Jan. 2012; Blockbuster, Sept. 2010; Netflix's $50 million offer, 2000. Pew: newspaper ads $49.4 billion (2005), $9.6 billion (2020); Medill 2025: ~3,500 papers closed since 2005. Yellow Pages Association: 14.5 billion lookups, 2005. Britannica: 120,000 sets in 1990; last print edition 2012. Nokia: sales €51.1 billion (2007), €19.9 billion (2025); value $156 billion (Oct. 2007) to $7.5 billion (June 2012), INSEAD. HP–Palm, 2010; Palm's IPO-close value, March 2, 2000: $53.4 billion (WSJ).

ABOUT ITERATE.AI

Iterate.ai is a private AI software infrastructure company based in San Jose, CA, and Colorado, building AI since 2015.

Iterate.ai — the private intelligence layer. Generate turns a company's own data into agents, answers, and decisions in minutes, inside the company's own infrastructure. AgentWatch provides observability and governance for AI agents. Lifeboat provides an inference runtime that runs models efficiently across different hardware. Iterate also builds and customizes private, industry-specific language models for individual organizations and industries.

Iterate.ai

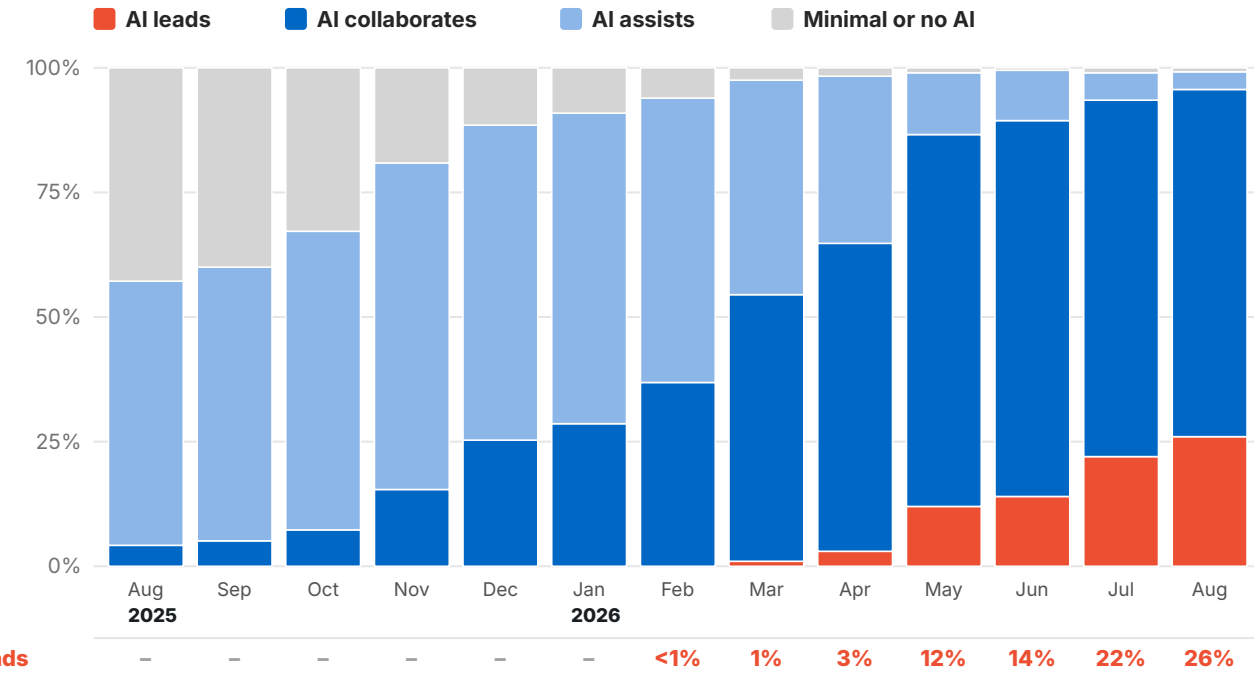
v14.6 · Nov 9, 2026 · Three addenda follow

ADDENDUM A · NOVEMBER 2, 2026

AI is now building AI

Page 8 described the third clock: each generation of AI helps build the next. In September 2026, Anthropic published the first public measurement of that loop at its own company. If that loop closes on its own, it is called recursive self-improvement. That is when AI builds the next AI on its own.

HOW MUCH OF ANTHROPIC'S AI RESEARCH WORK AI DOES, BY MONTH



	Aug 2025	Sep	Oct	Nov	Dec	Jan 2026	Feb	Mar	Apr	May	Jun	Jul	Aug
AI leads	–	–	–	–	–	–	<1%	1%	3%	12%	14%	22%	26%

Share of Anthropic's AI research and development work at each level of AI involvement. As of August 2026, 0% of the work is fully autonomous. The "AI leads" row is labeled in the chart as published; the other bands are read from it and are accurate to about one percentage point. Source: Anthropic, Sept 17, 2026, as shown by The Wall Street Journal, Oct 30, 2026.

What the levels mean

AI assists: AI helps, in a smaller role than collaborating.

AI collaborates: does large chunks of work under close human direction.

AI leads: completes most of a task from a high-level prompt while a person supervises.

AI is autonomous: no human in the loop. Anthropic says it has not reached this level.

What counts as research here

Every kind of work involved in building the next model. Anthropic sampled about 15,000 tasks from staff in July and sorted them into 378 kinds of work. Its examples include diagnosing evaluation-platform defects, setting network rules for training sandboxes, and writing postmortems on service incidents. One risk it names is AI choosing which research direction to pursue.

ADDENDUM A, CONTINUED · NOVEMBER 2, 2026

Agents at scale

~30,000
AI agents doing research and engineering work at once on one platform (Aug 2026)

None
of the measured work is fully autonomous

Above 90%
of the work is at "AI collaborates" or higher (Aug 2026)

Oversight at that scale

Anthropic says every action by those agents passes through an online monitor before it runs, usually within seconds, and is read again by an offline monitor afterward. Of more than a billion decisions in August, the online monitor blocked 0.002%, about 1 in 47,000. Anthropic's point is that individual agents rarely misbehave, but with millions or billions of agents, even rare events can happen regularly.

What to watch next

Other labs. Anthropic says any developer could publish the same measures today.

Outside checks. Anthropic plans to bring independent third-party evaluators inside the company.

The trend. Whether the "leads" share keeps rising, and whether AI starts choosing research direction.

What a leader can ask

Our work. How much of our own technical work does AI lead, and who measures it?

Our controls. What reviews an agent's action before it runs, not only after?

Our plan. Does it assume the speed of last year's AI?

WHY IT MATTERS, AND WHAT IT DOES NOT SHOW

If AI does more of the work of building AI, each generation can arrive sooner, and that speeds up the other three clocks. That is a possibility, not yet a measured trend. It is one company, six months of AI leading, and one method. Anthropic's own models did most of the rating. Their ratings matched staff exactly 59% of the time, while staff matched each other exactly 35% of the time, and the ratings were within one level of staff 97% of the time. Anthropic says it will keep publishing, so expect this page to age quickly.

SOURCES

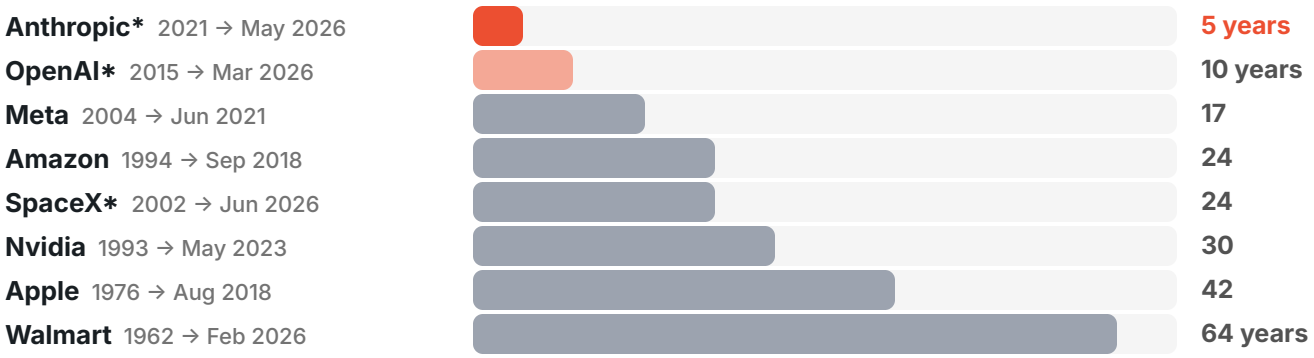
Anthropic, "Measurements for understanding the pace of AI development inside frontier labs," Sept 17, 2026: anthropic.com/institute/measuring-pace-of-ai-development. Automation scale: Epoch AI, epochai.substack.com.
Coverage and chart: Ben Cohen, "The Smartest People in AI Believe This Is the World's Most Important Chart," The Wall Street Journal, Oct 30, 2026.

ADDENDUM B · NOVEMBER 2, 2026

Five years to $965 billion

Apple became the first $1 trillion company in 2018; sixteen have crossed that line since, six of them since early 2025. Walmart took about 64 years from its 1962 founding, reaching it in February 2026. Apple took 42, Nvidia 30, Amazon 24, Meta 17. Anthropic, founded in 2021, was valued at $965 billion in May 2026 and filed to go public on October 26.

YEARS FROM FOUNDING TO A $1 TRILLION VALUATION



Years from founding to the first $1 trillion value, rounded. *Anthropic: private valuation, $965 billion (May 2026). OpenAI: private valuation, $852 billion (Mar 2026). SpaceX: value at its June 2026 IPO, $1.77 trillion.

Where it would rank (Nov 1 caps)

$965B, last round: about 16th.

$1.5T, reported low end: about 11th.

$2T, reported high end: about 7th.

$2.5T, illustrative: about 6th.

Companies it is valued above

At $965 billion it sits just below Walmart ($983B) and above JPMorgan Chase ($895B), AMD ($824B), Exxon Mobil ($764B), Johnson & Johnson ($658B), ASML ($604B), Costco ($449B), and Palantir ($442B). At $1.5 trillion it would also pass Walmart, Micron, Samsung, Berkshire Hathaway, and Eli Lilly.

WHAT THE COMPARISON DOES AND DOES NOT SHOW

Anthropic's number is a private funding-round valuation; the others are public market values that change daily. Its IPO has not priced. The $1.5 to $2 trillion range is a press report; talk slipped from about $2 trillion in August to $1.5 trillion in October as growth slowed. The filing shows a first-half 2026 net loss of about $80 billion, mostly a one-time charge, and a small operating profit in the second quarter. For leaders the point is speed: a five-year plan written in 2021 would have covered this company's whole life.

SOURCES

Anthropic Series G and H announcements; S-1, SEC EDGAR, Oct 26, 2026. IPO talk: Financial Times (Aug) and Reuters (Oct 15), via Leverage Shares; results via CNBC. Market caps: Forbes real-time data via WealthRank, Nov 1, 2026 (omits some companies, such as Saudi Aramco, so ranks could be one lower). $1 trillion dates and club size: CMC Markets, Admiral Markets, Kilocorn; Walmart: unanswered.io. OpenAI: Forge. SpaceX IPO: Wikipedia.

ADDENDUM C · NOVEMBER 2, 2026

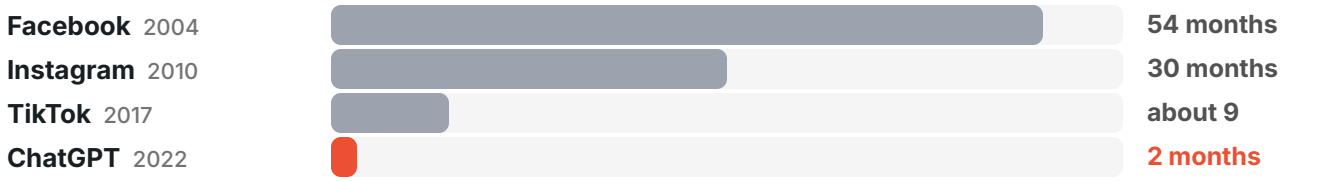
Is your inside clock as fast as the outside clock?

JACK WELCH, FORMER CEO, GENERAL ELECTRIC

> **"If the rate of change on the outside exceeds the rate of change on the inside, the end is near."**

This paper has measured the outside clock. This addendum asks about the inside one, starting with how fast people now adopt new technology.

MONTHS FROM LAUNCH TO 100 MILLION USERS



Facebook, Instagram: Visual Capitalist (PwC, Yahoo data). TikTok: about nine months after its global launch (Sensor Tower). ChatGPT: analyst estimate for January 2023 (UBS, Similarweb data); OpenAI reported 900 million weekly users by February 2026.

The outside clock

Models. Four frontier launches in eight days, July 2026 (page 8).

Agents and cost. Task length doubles every 4 to 7 months, and the cost of a given result falls about 13 times a year (page 10). Moore's law took about two years to double.

Research. AI leads 26% of Anthropic's AI research, from under 1% in February (Addendum A).

The inside clock: five things to time

From a new model release to your first test.

From a request for a new AI tool to a decision.

From a vendor change to your ability to swap.

Between board reviews of AI risk (about four meetings a year, page 18).

Between updates to your plan's assumptions.

WHAT THE COMPARISON DOES AND DOES NOT SHOW

The 100 million figures measure consumer apps, so the chart shows direction, not a law, and ChatGPT's two months is an analyst estimate. Matching the outside clock does not mean dropping control: where a mistake is costly, keep a person in the loop (pages 17 to 19 and 24).

SOURCES

Welch: attributed, wording varies (Raconteur; Tim Kastelle). Adoption: Visual Capitalist (PwC and Yahoo data); UBS and Similarweb via Reuters and The Guardian, Feb 2023; Sensor Tower via The Guardian; OpenAI via Reuters, Feb 2026. Moore's law: Gordon Moore, 1975. Other figures: pages 8 and 10 of this paper and Addendum A.