NEW HORIZON · AI DIGEST

The Long Read

7 November 2026

12	3	4154	3
STORIES	SECTIONS	WORDS	HIGH IMPACT



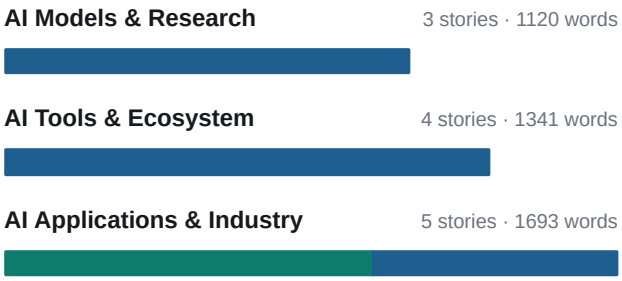
3 high impact 9 medium impact spoke length = depth of coverage

The full-length companion to the daily New Horizon AI Digest.
Every story in today's email, reported at length · new-horizon.tech/digest

The issue at a glance

12 stories · 4154 words · 3 sections · 4 charted





AI Models & Research

01 Researchers Use AI to Compare Competing Scientific Models

02 Anthropic Finds Better Agent Memory Means Writing Less

03 ToyBlocks Exposes Robots' Weakness on Unfamiliar Tasks

AI Tools & Ecosystem

04 Meta Gives Muse Persistent Spaces for Projects

05 Amazon Lets Alexa Handle Shopping by Text Message

06 Underdog Shrinks Offline Speech Generation to 32MB

07 Speech Tests Add Egyptian and Nigerian English

AI Applications & Industry

08 Google Takes Over Crusoe's Unfinished Wyoming Data Center

09 AI Spending Barely Grew as Price Cuts Took Hold

10 Hundreds of AI Hacking Systems Were Left Open Online

11 Chainalysis Breach Exposed Credentials for Every Customer

12 Tau's Home-Cleaning Robots Still Depend on Remote Humans

How to read this. Every story in today's email is reported here at full length, in the same order. Impact is the writer's judgement of whether a story changes what a reader — in business, at work, as an investor or as a builder — should do or believe this week. Charts appear only where the source itself puts comparable numbers side by side; nothing is estimated to fill a gap. Sources are listed in full on the final page.

SECTION 1 OF 3

AI Models & Research

3 stories ● 3 medium

01 MEDIUM IMPACT www.anthropic.com

Researchers Use AI to Compare Competing Scientific Models

> Researchers at the ELLIS Institute Tübingen and the University of Tübingen say AI can speed up scientific modelling, but only if it shows the range of plausible explanations rather than handing over a single model.

Jakob Macke, Daniel Gedon and Stefan Wahl, writing in a guest post for Anthropic, set out how AI is changing three parts of scientific work: building a model, fitting it to measurements and interpreting what it says about the real world. Their central warning is that a model which reproduces the data is not necessarily an explanation. Many different models can fit the same results, and an AI agent that returns just one hides the alternatives.

The more established advance is in fitting models. A method called simulation-based inference runs a computer model repeatedly with different settings, then trains a neural network to work backwards from an observed result to the range of settings that could have produced it. The simulations can take weeks, but once trained the network can analyse new data in seconds. The approach has been used in cosmology, structural biology and geophysics; when two neutron stars collide, it can estimate their location quickly enough for telescopes to be pointed while the signal is still detectable.

The newer step is letting AI design the model itself. Systems appearing over the past two years use a large language model to write candidate models as code, test them against data and revise them. The researchers argue that these systems should also produce a spread of plausible models, with weaker candidates discarded and new variations built on stronger ones. Features shared by all the survivors are likely to matter; disagreements show where another experiment is needed. But the evidence is thin where it counts most: these systems have mostly been tested on rediscovering models scientists already know, with sparse evidence on unanswered problems.

The historical comparison makes the stakes clear. In 1951, Andrew Huxley spent more than three weeks solving nerve-cell equations on a hand-cranked calculator after Cambridge University's computer went out of action for six months. An AI system could plausibly find a similar equation in an afternoon. Yet Hodgkin and Huxley's crucial insight came from knowing that simpler versions failed: their fourth-power term suggested four moving parts in a cell-membrane gate, a hypothesis later supported by protein structure. Interpretation remains the hardest step to automate, because there is no clear way to measure whether an AI system understands a system rather than merely fits its data.

KEY FACTS

RESEARCHERS

Jakob Macke, Daniel Gedon and Stefan Wahl, at the ELLIS Institute Tübingen and University of Tübingen

SIMULATION WORKLOAD

Perhaps 100,000 runs for a Hodgkin-Huxley-style model, and millions for larger models

EVIDENCE SO FAR

Model-building systems have mostly rediscovered known models; evidence on unsolved problems is sparse

HISTORICAL COMPARISON

Huxley took over three weeks with a hand-cranked calculator; AI could plausibly find a similar equation in an afternoon

SPEED AFTER TRAINING

Inference on new data takes seconds, without further simulations

> WHY IT MATTERS
>
> For organisations using AI in research, the practical lesson is to ask for competing explanations and unresolved assumptions, not just a model that fits. Treat AI-generated scientific answers as a guide to further testing, especially since success on genuinely unsolved problems is still sparse.

Read the original at www.anthropic.com →

---

02 MEDIUM IMPACT www.anthropic.com

Anthropic Finds Better Agent Memory Means Writing Less

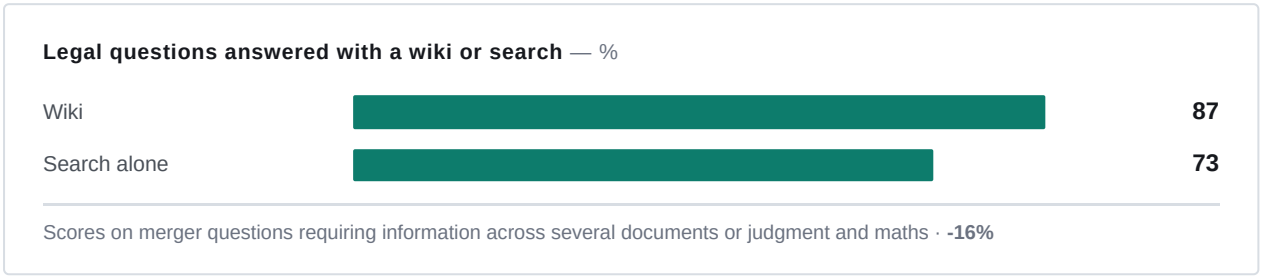
> Anthropic says its latest AI agents work better with smaller, carefully maintained memories, and that reviewing old session logs can correct mistakes that would otherwise persist.

Anthropic has published research on how its Claude agents retain knowledge between sessions, using notes, searchable records and a process it calls dreaming. During dreaming, the agent uses idle computing time to review earlier work, check notes against original transcripts and reorganise what it has learned. The findings matter to businesses deploying agents for long tasks: persistent memory can help, but indiscriminate note-taking consumes time and can preserve errors.

The clearest example came from Pokémon Red, which Anthropic used to test judgment over extended tasks. About 10,000 steps into separate runs, Claude Sonnet 3.7 had written 413 memory files across 225 folders and earned one badge. Claude Fable 5 had written nine files in two folders, plus one overview file, and earned four badges. The older model recorded battles, items and other details that were often visible on screen or quickly became outdated. At one point it spent 17 steps creating and deleting memory files without taking any action in the game. Fable 5 instead kept plans, strategies and tested rules. Anthropic says Opus 5.5 later used a similarly small set of notes and beat the game.

The research also challenges a natural instinct for companies building with agents: prescribing exactly what they should remember. On a database-question benchmark, telling Opus 5 and Opus 4.7 to record only failures removed much of memory’s performance benefit. Newer models generally performed better when allowed to judge what was worth saving. Older models needed more direction; instructing Sonnet 4.6 to read its notes and test its theories nearly doubled its benchmark score.

Dreaming supplied a further improvement by auditing memory after a session. In one Pokémon run, Claude recorded a landmark on the wrong side of a mountain and repeatedly took a wrong turn. A dreaming pass compared the note with the transcript, corrected it, and the agent subsequently reached its destination. Anthropic also tested agent-built wikis on legal and company documents. They helped when answers required information from several documents, but offered no clear gain when a simple text search sufficed; on one task, the wiki grew larger than the underlying documents. The genuinely new contribution is this set of evaluations and design lessons, rather than a new memory product. Anthropic says open questions remain about computing costs, update frequency and preventing accumulated errors.



KEY FACTS

POKÉMON MEMORY COMPARISON
Sonnet 3.7 wrote 413 files and earned one badge; Fable 5 wrote nine files in two folders plus an overview and earned four

COMPANY-DOCUMENT TEST
About 350,000 Stainless files, evaluated with 25 questions

WIKI-BUILDING RESULTS
Performance rose from 13% to 48% of graded criteria as dreaming increased from two sessions to 128

MULTI-DOCUMENT LEGAL QUESTIONS
Claude scored 87% using a wiki versus 73% using search alone

CONTRACT-BRIEF IMPROVEMENT
Revised wiki instructions raised met criteria from 72% to 100%

CLAUDE CODE MEMORY LIMIT
Only the first 200 lines or 25 KB of MEMORY.md are loaded into each session

WHY IT MATTERS

Businesses building or buying AI agents should treat memory as something to curate and audit, not simply maximise. Better results may come from concise notes, freedom for stronger models to judge importance, and scheduled reviews that catch persistent mistakes.

Read the original at www.anthropic.com →

03 MEDIUM IMPACT huggingface.co

ToyBlocks Exposes Robots’ Weakness on Unfamiliar Tasks

> Allen AI has released ToyBlocks, an open robot-testing world whose initial experiments expose weaknesses in how today’s models handle unfamiliar tasks and physical objects.

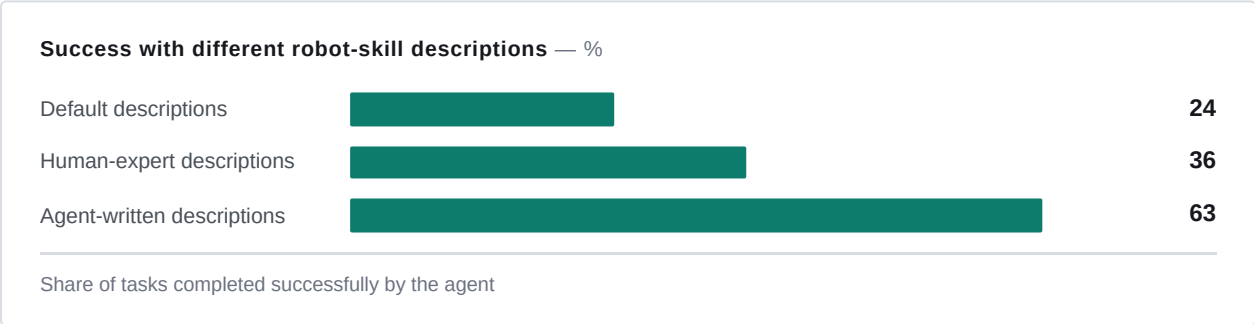
ToyBlocks is a simulated tabletop with a robot arm and lettered blocks. Researchers can automatically generate tasks and training demonstrations, then test skills such as pushing, flipping and stacking. It is the first of Allen AI’s planned FoundationWorlds: controlled settings intended to make robot experiments repeatable and failures easier to trace. The release includes simulation tools, training and evaluation code, and two trained models.

The early results are a warning against reading polished demonstrations as evidence of broad competence. Initial tests of ManiFlow, π0.5 and MolmoAct 2 found that these action models generalised poorly despite large-scale pretraining. The experiments varied block positions, stack heights, letters and colours, including tasks outside the training range. Allen AI says action-model generalisation remains far from solved. That matters to companies developing warehouse, manufacturing or domestic robots: performance in a familiar setup may not predict what happens when the layout or task changes.

The most striking finding concerns a Claude-based agent asked to build towers spelling three-letter words. Half the scenes lacked a required letter, so the correct response was sometimes to declare the task impossible. It solved the task or correctly reported infeasibility in 63% of cases; errors included confusing look-alike letters and choosing the wrong orientation. When the agent ran small experiments and rewrote the descriptions of its own skills, success rose from 24% with default descriptions to 63%, beating 36% with descriptions written by a human expert.

The framework also produced evidence that simulation can transfer to physical machines. ManiFlow's individual skills succeeded 80.0% of the time in simulation and 81.1% on a real robot arm. A combined Molmo 2 and ManiFlow system completed 50.0% of real-world block-spelling episodes, against 55.5% in simulation, though Allen AI says more trials are needed. Separately, adding 100 trajectories generated by the best-performing video model raised a policy's success in an unseen region from 19% to 78%.

The genuinely new contribution is the controlled, open testing and data-generation system, rather than a claim of a generally capable robot. The initial release contains 85,336 long-horizon demonstrations, covering nearly 1,600 hours of robot execution. A larger dataset, more skills and broader benchmarking are targeted for the coming weeks.



KEY FACTS

WHAT WAS RELEASED
ToyBlocks, an open simulation framework for robot learning and evaluation

INITIAL TRAINING DATA
85,336 long-horizon demonstrations, covering nearly 1,600 hours of robot execution

AGENT SUCCESS AFTER WRITING ITS OWN SKILL DESCRIPTIONS
63%, up from 24% with default descriptions and 36% with human-expert descriptions

INDIVIDUAL ROBOT SKILLS
80.0% success in simulation and 81.1% in the real world

REAL-WORLD BLOCK-SPELLING TASKS
50.0% success, or 15 of 30 episodes

NEXT RELEASE
A larger dataset, more skills and broader benchmarking targeted within a few weeks

WHY IT MATTERS

Businesses assessing robot AI should demand tests on unfamiliar layouts and tasks, not just polished demonstrations. For researchers and developers, ToyBlocks offers a reusable way to find failures and measure whether extra training data actually improves performance.

Read the original at huggingface.co →

SECTION 2 OF 3

AI Tools & Ecosystem

4 stories ● 4 medium

04 MEDIUM IMPACT TechCrunch

Meta Gives Muse Persistent Spaces for Projects

> Meta has launched Spaces, a feature that gives its Muse agents persistent project areas where users can return to ongoing tasks and connect outside apps for automatic updates.

Spaces are breakout conversations organised around a particular topic or goal, living in the sidebar of the interface where users interact with their Muse agent. Clicking on a Space opens the conversation, where Muse works toward that goal on the user's behalf. Meta describes them as "custom experiences" and says users can describe an idea in chat and Muse will build a space for it, whether that is a tracker, a planner or a project board. Examples the company gives include planning a vacation, collecting recipes for a family event, tracking holiday gifts, planning a wedding or managing DIY projects.

The more consequential addition is integration with third-party apps and information. Meta says that once users connect the apps they use, their spaces keep themselves up to date in the background, so new concert dates, price changes or savings progress appear without the user having to ask. That moves Muse beyond a chat window that waits for instructions and toward something that monitors connected services and maintains a live view of a project. For businesses building consumer apps or services, it is a signal that Meta wants its agent interface to sit between users and the apps they already rely on.

The feature is the second Muse addition Meta has announced in the past week. On Thursday, the company introduced Muse Voice, which lets users video chat with their agent and give it commands verbally. Mark Zuckerberg, Meta's founder and chief executive, posted on X on Friday that he had been playing with spaces in Muse and had made a barbecue game, an event tracker and a home control dashboard.

Meta is building on early traction. During its earnings call in October, Zuckerberg said the agents had reached 10 million users in the United States and Canada, less than two months after they debuted. The company is competing with other cartoon-styled AI agents, including OpenAI's dots and SpaceXAI's Grokbot, which launched with similarly playful personas. Spaces itself is an incremental feature — persistent project threads are not a new idea — but combined with app connections and voice, it shows Meta moving quickly to turn an early user base into a more capable assistant.

KEY FACTS

MUSE USERS
10 million in the US and Canada, less than two months after launch

WHAT SPACES DOES
Persistent project conversations in the Muse sidebar, with third-party app connections that update in the background

ALSO LAUNCHED THIS WEEK
Muse Voice, for video chat and verbal commands

COMPETITORS NAMED
OpenAI's dots and SpaceXAI's Grokbot

> WHY IT MATTERS
>
> For businesses and developers, Meta's push to connect Muse with third-party apps could create a new channel through which customers interact with their services. For investors, the 10 million user figure offers an early measure of whether Meta's agent bet is gaining traction against rivals.

Read the original at TechCrunch →

---

05 MEDIUM IMPACT TechCrunch

Amazon Lets Alexa Handle Shopping by Text Message

> Amazon is launching a text-message shopping service with Alexa for select US customers, just before the holiday shopping season.

Amazon said on Friday that customers invited to the service will receive a text asking them to enable it. They can then message Alexa in ordinary language to find products, add them to a cart and complete a purchase on Amazon. Others can join a waitlist, with a broader rollout planned in the coming weeks.

The service extends Alexa's recent shopping upgrades rather than starting from scratch. It can manage lists, track prices, search for items across the web, compare products and check order status. Shoppers can also send a picture of an item, or of someone wearing an outfit, to find similar listings on Amazon. The company says Alexa will learn users' preferences over time and suggest products accordingly.

The timing exposes the commercial tension behind consumer AI agents. Amazon blocked Meta's Muse from querying its site in September; it is now offering its own text-based route into shopping, similar to agents such as Muse and Instinct. For retailers and brands, that raises the possibility that an AI assistant controlled by the platform becomes another gatekeeper between shoppers and product listings. For Amazon, it creates a direct way to keep product searches, price comparisons and purchases inside its own service.

The most striking evidence that this market is still early comes from DoorDash. The delivery company said last month that third-party agents had placed only hundreds of orders in total on its platform so far. It plans to concentrate on its own text-based agent, which launched in September. That is a small base compared with the attention consumer agents are attracting, and it suggests the immediate contest is over who owns the customer relationship, not over a proven surge in agent-driven sales.

KEY FACTS

INITIAL AVAILABILITY
Select US customers, with a waitlist for others

WIDER ROLLOUT
Coming weeks

WHAT SHOPPERS CAN DO
Search, add items to cart and purchase by texting Alexa

EARLIER RESTRICTION
Amazon blocked Meta's Muse from querying its site in September

DOORDASH THIRD-PARTY AGENT ORDERS
Hundreds of orders in total so far

> WHY IT MATTERS
>
> Amazon is testing whether its own assistant can become the front door to shopping after shutting out an outside agent. Retailers and brands should watch whether product discovery and purchasing shift further into platform-controlled AI services, though DoorDash's figures show agent-driven orders remain small.

Read the original at TechCrunch →

---

06 MEDIUM IMPACT TechCrunch

Underdog Shrinks Offline Speech Generation to 32MB

> AI startup Underdog has released Woof, a speech-generation model small enough to run offline on everyday devices, and demonstrated it through a browser-based game.

The game presents a retro-themed room containing a typewriter, a CRT monitor, a desk lamp and a music tuner. Users can type text on the typewriter and have it spoken back, or rotate the tuner to generate different kinds of music. The lamp also makes a sound when its switch is clicked. An introductory warning about “intrusive” cloud providers in the town is an apparent dig at cloud-based AI models.

The demonstration is powered by Woof, which Underdog says can run on most devices without an internet connection. The model is 32MB in size, has 31 million parameters and is open for anyone to download. Parameters are the internal settings that determine how an AI model behaves; the important point for non-specialists is that this system is designed for local use rather than relying on remote servers.

That could make the technology useful for local assistant experiences or personalised podcasts, although the source cautions that such a small model is unlikely to replace larger systems. For businesses and developers, the attraction is the possibility of adding speech features without depending on a cloud provider. For users, running the model on their own device offers a route to keeping their data under their own control.

Woof follows Bark, a 167MB speech-to-text model that Underdog launched weeks earlier for transcription and dictation on edge devices such as phones and computers. Founder Sigil Wen told TechCrunch that he began building the app because he wanted an alternative to notetakers such as Granola, using AI models that run on users’ devices and let them own their data. He said he expected “most capable models of today” to run on devices people already own within months, and described Underdog’s goal as training capable, open, American-made models that enable private AI.

The company has also released a 27-billion-parameter model that can run on laptops. Separately, it has a waitlisted app positioned as a locally hosted alternative to Muse and Instinct. The genuinely new element here

is the extremely small offline speech generator and its interactive browser demonstration; the broader effort to move AI processing from the cloud onto personal devices is already being pursued by other researchers and startups.

KEY FACTS

NEW SPEECH MODEL
Woof, a 32MB model with 31 million parameters

HOW IT RUNS
On most devices without an internet connection

AVAILABILITY
Open for everyone to download

EARLIER MODEL
Bark, a 167MB speech-to-text model launched weeks earlier

OTHER RELEASE
A 27-billion-parameter model that can run on laptops

> **WHY IT MATTERS**
>
> For businesses and developers, Woof points to speech features that can run on customers' own devices without an internet connection or cloud processing. That could support more private assistants, transcription tools and audio products, though the small model is not expected to replace larger systems.

Read the original at TechCrunch →

---

07 MEDIUM IMPACT huggingface.co

Speech Tests Add Egyptian and Nigerian English

> Hugging Face has added Egyptian and Nigerian English to the private tests behind its Open ASR Leaderboard, which ranks speech-recognition models.

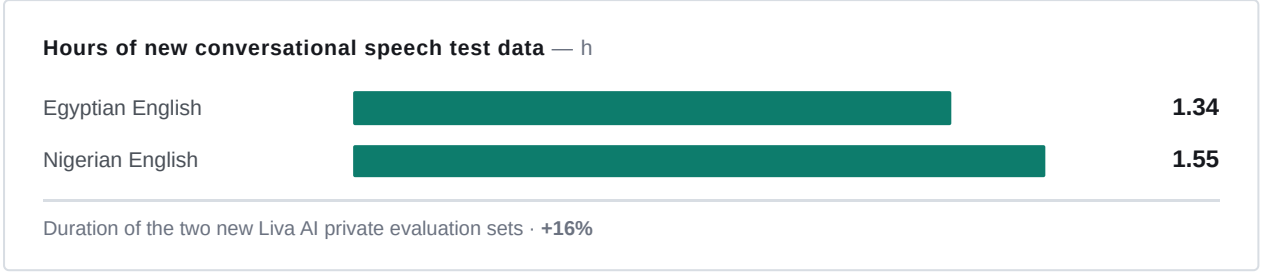
The new material comes from Liva AI and covers spontaneous, conversational speech rather than scripted readings. The Egyptian set runs to 1.34 hours and the Nigerian set to 1.55 hours, with transcripts that keep punctuation, capital letters and disfluencies — the hesitations and false starts of ordinary talk. Both are private evaluation sets: the audio is not published for model makers to train on.

That privacy is the point. A few months ago the leaderboard added 11 private splits from Appen and DataOceanAI covering scripted and conversational speech in five accents, and it has since added an Indian-accented English split with Voice Arena. The organisers say the new data is meant to minimise "benchmaxxing" — tuning a system to do well on a known test rather than on speech it has never seen. For a business buying transcription or voice-agent software, a ranking built on held-out recordings is a more useful signal than one a vendor could rehearse for.

The leaderboard is also changing how it reports results, grouping accents using Kachru's Three Circles, a framework credited to Hume AI's leaderboard, and promising to split the Outer and Expanding Circles as more accents arrive. The aim is a breakdown that shows where a model works and where it does not, instead of a single English score that can hide weak performance on speakers outside the US and Britain.

The genuinely new part is narrow: two accents, under three hours of audio between them, and one more data provider. The incremental part is the continuation of a stated push for more accents, more languages and more

private providers. Liva AI says it collects multimodal interaction data through consumer products used by real people, and that owning both collection and distribution lets it build products around specific model gaps, yielding natural behaviour rather than task-based interactions. No model rankings or accuracy figures were published with the announcement, so there is no evidence yet of which systems gain or lose from the tougher tests.



KEY FACTS

NEW TEST ACCENTS
Egyptian and Nigerian English

NEW AUDIO
1.34 hours Egyptian; 1.55 hours Nigerian, conversational and spontaneous

DATA PARTNER
Liva AI

EARLIER PRIVATE TESTS
11 splits from Appen and DataOceanAI across five accents, plus an Indian-accented split with Voice Arena

WHY THE DATA STAYS PRIVATE
To minimise benchmaxxing — models being tuned to known benchmark material

> **WHY IT MATTERS**
>
> If you buy or build speech-recognition tools, rankings based on private, accented speech are harder to game and more likely to reflect real customers. The Egyptian and Nigerian additions are small, so treat this as a credibility improvement rather than proof that any model now handles those accents well.

Read the original at huggingface.co →

SECTION 3 OF 3

AI Applications & Industry

5 stories · 3 high · 2 medium

08 MEDIUM IMPACT TechCrunch

Google Takes Over Crusoe’s Unfinished Wyoming Data Center

> Google is taking over an unfinished data center project in southeastern Wyoming from Crusoe, the developer that had tried and failed to sign it up as a customer.

Crusoe confirmed on Friday that it had sold the project to Google. The search company had reportedly raised concerns about the development’s cost and construction timeline under Crusoe’s management. Google did not disclose the purchase price, but said it planned to invest $17 billion in developing the data center campus.

The project was announced last year by Crusoe and Tallgrass, an infrastructure company backed by Blackstone, with plans eventually to reach 10 gigawatts of power capacity. Its change of ownership shows the difficulty independent developers face in delivering very large data centers on time and within budget. Securing equipment and power is challenging, and opposition from local communities can add further complications. For businesses buying AI computing capacity, the deal is a reminder that announced campuses do not necessarily become available on the timetable first promised.

The most striking detail is that Google has moved from prospective customer to owner after Crusoe failed to land its business. That is a setback for Crusoe’s role on this project, but not a complete break between the companies: Crusoe is still developing another data center for Google near Amarillo, Texas. Bloomberg first reported the Wyoming sale in June, and Crusoe went on to raise $3.9 billion in September despite the setback.

Crusoe, an eight-year-old Denver-based company recently valued at $30.9 billion, leases data center space to customers that bring their own AI chips, rents out its own chips and sells computing power for running AI models. Its largest projects serve major technology groups. Its Abilene, Texas campus is being built in phases for Oracle and OpenAI, while a separate 900-megawatt campus next door is being constructed for Microsoft, with the first building expected to come online in mid-2027. Meta is also a customer, and Bloomberg reported that trading firm Jane Street signed a $13 billion, five-year cloud deal with Crusoe. The genuinely new development is Google’s takeover of the Wyoming site; Crusoe’s broader business and other projects remain in place.

KEY FACTS

GOOGLE'S PLANNED CAMPUS INVESTMENT
$17 billion

PURCHASE PRICE
Not disclosed

CRUSOE'S RECENT VALUATION
$30.9 billion

CRUSOE'S SEPTEMBER FUNDING ROUND
$3.9 billion

PLANNED WYOMING POWER CAPACITY
10 gigawatts eventually

FIRST MICROSOFT BUILDING
Expected online in mid-2027

WHY IT MATTERS

The takeover shows that even well-funded data center developers can struggle to deliver huge AI projects on budget and on schedule. Businesses and investors should distinguish announced capacity from capacity that is actually built and available.

Read the original at TechCrunch →

09 HIGH IMPACT TechCrunch

AI Spending Barely Grew as Price Cuts Took Hold

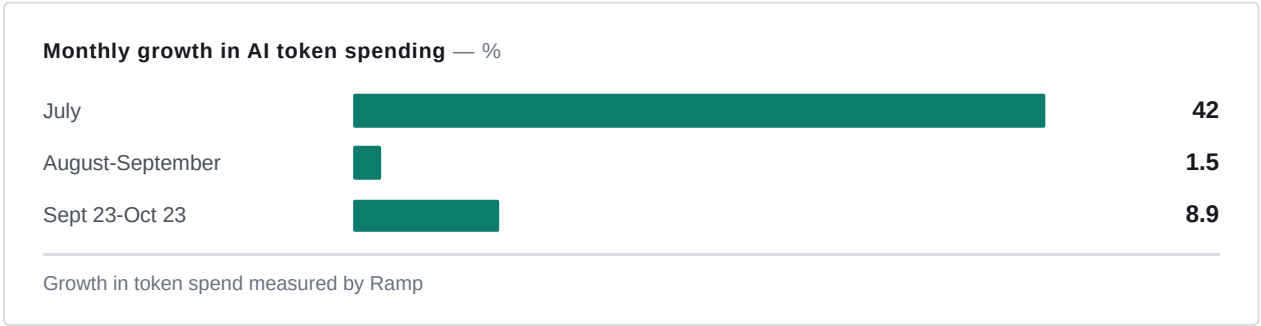
> Companies are using more AI than ever, but spending on it has stalled as price cuts and savvier customers squeeze model makers' revenue.

A new report from Ramp shows AI spending flatlined over late summer even as usage volumes rose. Token spend grew just 1.5% across August and September, compared with 42% growth in July from the previous month. October brought a partial rebound, with token spend up 8.9% from September 23 to October 23, but Ramp's lead economist Ara Kharazian said the headwinds on spending growth remain strong.

The slowdown is not a demand problem. Kharazian points instead to price competition: OpenAI cut API prices in half for GPT-6 Sol and GPT-6 Luna in September, and Anthropic made significant reductions to pricing for Fable 5.1 and Opus 5.5, a decision the company said in a recent filing slashed revenue growth by about $600 million. Both companies have also offered enterprise discounts intended to lock in spending that are instead hitting their revenue targets. Customers are getting thriftier too, setting budgets for employees and routing work to cheaper models. Notion co-founder Akshay Kothari posted internal data matching Ramp's findings, summarised as: "tokens go brrr. costs go ahhh."

The concentration of spending makes the slowdown sharper. Almost all AI spending comes from venture-backed, high-growth tech companies, and the top 1% of firms have stuttered slightly, falling just below the $10,000 per employee they were hitting at summer peaks. Median firms spend $12 to $15 per employee per month, and the top 10% spend $600 to $1,000. Kharazian said a slowdown among that top cohort is sufficient to slow revenue entirely despite rising volumes, and it is not clear other industries will ever reach that level: a doctor's office or car dealership is unlikely to need to spend that much.

The timing is awkward for Anthropic, which is gearing up for an IPO. Bloomberg reported Friday that its potential valuation sank from the $2 trillion estimated in early October to $1.2 to $1.3 trillion, with a high junk rating expected from credit agencies. OpenAI lowered its proposed valuation to $1.25 trillion from $1.4 trillion and cut its latest fundraising goals by about $10 billion. Investors hope new products such as personal agents will drive the next wave of spending, but Kharazian is skeptical they are as economically productive as coding agents, which emerged just a year and a half ago as the most commercially viable AI product.



KEY FACTS

AI SPENDING GROWTH, AUGUST-SEPTEMBER
1.5%

AI SPENDING GROWTH IN JULY
42% from the previous month

ANTHROPIC REVENUE GROWTH HIT FROM PRICE CUTS
About $600 million

TOP 1% OF FIRMS' AI SPEND
Just below $10,000 per employee at summer peaks

MEDIAN FIRMS' AI SPEND
$12-$15 per employee per month

ANTHROPIC POTENTIAL VALUATION
$1.2-$1.3 trillion, down from $2 trillion in early October

> WHY IT MATTERS
>
> Businesses buying AI now have leverage: price cuts and competition mean the same workloads cost less. Investors and model companies face a harder question, since rising usage no longer guarantees rising revenue.

Read the original at TechCrunch →

---

10 HIGH IMPACT TechCrunch

Hundreds of AI Hacking Systems Were Left Open Online

> Security researchers have found more than 800 copies of a Chinese-made AI hacking tool sitting openly on the internet, with no password needed to log in.

Censys, a cybersecurity firm that scans the internet for exposed systems, identified the deployments of Artex, an open source tool that automatically probes computer systems for weaknesses. Artex has been linked to several cyberattacks since its public release in July. Microsoft has described it as a system of AI agents that can conduct attacks from reconnaissance through data theft with little human interaction. Because it is free to download and run, it has become popular with people who cannot access Anthropic or OpenAI's most advanced cybersecurity models.

The exposure cuts both ways. An open Artex system could let an outsider use it to attack unsuspecting targets, but it also reveals the operator's own attack tools, including implants — hidden malicious software used to keep access to a victim's network. Silas Cutler, the Censys principal security researcher who wrote the findings, said not every instance is run maliciously and many are used by authorized testers. It is unclear why so many are reachable from the web, since the software does not need to be.

The linked attacks show the stakes for ordinary businesses. Security companies have tied Artex to a September campaign against the Korean finance industry, which CrowdStrike attributed to a Chinese-speaking hacker who had left an open directory hosting the system. Censys said that instance had been publicly exposed for months and appeared to hold attack data reaching well beyond South Korean banks, including payment and gaming companies. Artex was also used in an October attack on IT support giant ConnectWise that let Chinese-linked hackers spy on thousands of organizations whose data and files are hosted in Microsoft's cloud. Microsoft said the hackers broke into customers' cloud environments and stole data so quickly that defenders had little time to detect and respond.

The genuinely new finding is the scale of the exposure, not the existence of AI-assisted hacking. Many of the attacks linked to Artex relied on basic weaknesses such as weak or known passwords and a lack of multi-factor

authentication, the common login safeguard that asks for a second proof of identity. Artex has since said it will go closed-source, limiting further releases to paid customers, but dozens of copies of its code remain on code-sharing sites and can still be used in active attacks.

KEY FACTS

OPEN ARTEX SYSTEMS FOUND
More than 800, accessible without a password

TOOL RELEASED
July, as free open source software

KOREAN FINANCE ATTACKS
September; linked instance exposed for months

CONNECTWISE ATTACK
October; thousands of organizations spied on

CODE STILL CIRCULATING
Dozens of copies on code-sharing sites

> **WHY IT MATTERS**
>
> For any business, the immediate lesson is unglamorous: weak passwords, missing multi-factor authentication and misconfigured systems are what these AI-assisted attacks keep exploiting. The exposed Artex servers also mean a free attack tool and its operators' own data are both sitting within reach of anyone who finds them.

Read the original at TechCrunch →

---

11 HIGH IMPACT Ars Technica

Chainalysis Breach Exposed Credentials for Every Customer

> Blockchain analysis firm Chainalysis says hackers obtained sensitive login data for all of its users after breaching its network security.

The stolen information includes names, usernames, password hashes and live tokens. Password hashes are scrambled versions of passwords, while live tokens can act like temporary digital keys for an account. Chainalysis said the attackers gained unauthorised access to its systems and internal tools using AI agent-assisted offensive techniques.

The customer list makes the breach especially sensitive. Chainalysis and public sources have identified users including the FBI, US Immigration and Customs Enforcement, the US Air Force, Microsoft, Coinbase and Barclays. The company analyses blockchain activity, so its tools sit close to investigations and financial compliance work. It remains unclear whether the theft exposed customers to further threats.

The company’s account of what was taken has also changed. In an email last Friday, Chainalysis told affected customers that an October cybersecurity incident had exposed what it called “non-sensitive customer data”, including user IDs, full names, email addresses and organisation role labels. The source questions why that information was treated as non-sensitive. In a Wednesday update, Chainalysis said the incident was part of an “ongoing multifront cyberattack” and that a separate on-path attack had exposed a subset of customer passwords.

An on-path attack, also called an adversary-in-the-middle attack, involves intruding into infrastructure that carries traffic between two parties. Chainalysis said that attack route had been identified and fixed, but it has

not said where the infrastructure was compromised or whether the captured data was protected by TLS encryption, a standard method for securing data in transit. The genuinely new development is the Wednesday disclosure of password exposure and the claim that credentials and tokens cover all users, beyond the more limited customer-data access described earlier.

KEY FACTS

USERS AFFECTED
All Chainalysis users

DATA OBTAINED
Names, usernames, password hashes and live tokens

INCIDENT TIMING
Cybersecurity incident in October, disclosed last Friday

ATTACK METHOD
AI agent-assisted offensive techniques

NAMED CUSTOMERS
FBI, US Immigration and Customs Enforcement, US Air Force, Microsoft, Coinbase and Barclays

PASSWORD EXPOSURE
A subset exposed through an on-path attack; the attack vector was identified and fixed

> **WHY IT MATTERS**
>
> Any organisation using Chainalysis should treat this as a credential-security incident, not just a privacy notice. The use of AI-assisted attack techniques against a firm serving law enforcement and financial companies also shows the practical security risk facing businesses that rely on sensitive third-party platforms.

Read the original at Ars Technica →

---

12 **MEDIUM IMPACT** MIT Technology Review

Tau’s Home-Cleaning Robots Still Depend on Remote Humans

> Tau Robotics has begun cleaning San Francisco homes with humanoid robots that still depend heavily on remote human operators.

The startup’s robots can vacuum and mop without human help, according to cofounder Cornelia Weinzierl, but tasks such as making beds and folding laundry still require a person to guide their movements remotely. Pilots watch through cameras in the robots’ heads and wrists. For some jobs, the operator performs the motions and the robot copies them, generating examples that Tau hopes will teach future machines to work independently.

That makes the service less a finished product than a way to collect training data from real homes. Tau has completed only a couple hundred cleanings so far. Its pricing reflects the experiment: an early offer was $30 for an hour with one robot, while the current charge is $60 for a weekly cleaning with two robots or $100 for a one-time visit. The reporter who tested it found the machines thorough but slow, and concluded that cleaning yourself would be a better use of time than watching them work.

The visit also showed how far the machines are from the effortless domestic robots of science fiction. The four-and-a-half-foot-tall machine staggered up the front steps, was defeated by a carpeted staircase and had to be carried. It later stopped cleaning when hot weather disabled the equipment used by its human controller. On a smoother second visit, it cleaned the kitchen and toilet, but left taps dripping, knocked over a closed olive-oil

bottle and could not right an air purifier it had toppled. Its most impressive trick was repeatedly reattaching the head of a toilet-cleaning wand after it fell off.

The genuinely new part is that a startup is putting humanoid cleaners into ordinary homes and charging customers, rather than demonstrating them in a laboratory. The autonomy remains incremental. If Tau or its competitors eventually make cleaning cheap and independent of remote pilots, professional cleaners could lose work. Weinzierl argues that robots could fill jobs marked by high turnover and injury risk, and says Tau is considering other demanding uses such as mold remediation and home health care.

KEY FACTS

CURRENT WEEKLY PRICE
$60 per cleaning with two robots

ONE-TIME VISIT
$100

CLEANINGS COMPLETED
A couple hundred

ROBOT HEIGHT
Four and a half feet

TASKS ALREADY UNAIDED
Vacuuming and mopping

> **WHY IT MATTERS**
>
> For businesses and investors, Tau shows that paid home robotics can arrive before full autonomy, with customers effectively subsidising data collection. For cleaners, widespread job displacement remains a possibility rather than an immediate consequence of a service that has completed only a couple hundred visits.

Read the original at MIT Technology Review →

Sources

01 Researchers Use AI to Compare Competing Scientific Models
https://www.anthropic.com/research/teaching-ai-to-build-scientific-models

02 Anthropic Finds Better Agent Memory Means Writing Less
https://www.anthropic.com/research/components-of-agent-memory

03 ToyBlocks Exposes Robots' Weakness on Unfamiliar Tasks
https://huggingface.co/blog/allenai/toyblocks

04 Meta Gives Muse Persistent Spaces for Projects
https://techcrunch.com/2026/11/06/meta-introduces-spaces-a-new-way-to-collaborate-with-your-muse-agent/

05 Amazon Lets Alexa Handle Shopping by Text Message
https://techcrunch.com/2026/11/06/after-blocking-muse-amazon-now-lets-users-text-alexa-to-help-them-shop/

06 Underdog Shrinks Offline Speech Generation to 32MB
https://techcrunch.com/2026/11/06/underdog-shows-off-a-32mb-speech-model-in-new-demo/

07 Speech Tests Add Egyptian and Nigerian English
https://huggingface.co/blog/open-asr-leaderboard-liva

08 Google Takes Over Crusoe's Unfinished Wyoming Data Center
https://techcrunch.com/2026/11/06/crusoe-sells-troubled-wyoming-data-center-project-to-google/

09 AI Spending Barely Grew as Price Cuts Took Hold
https://techcrunch.com/2026/11/06/ramp-data-shows-ai-spending-stalling-just-as-anthropic-nears-its-ipo/

10 Hundreds of AI Hacking Systems Were Left Open Online
https://techcrunch.com/2026/11/06/hundreds-of-artex-ai-hacking-tools-found-publicly-exposed-to-the-web/

11 Chainalysis Breach Exposed Credentials for Every Customer
https://arstechnica.com/security/2026/11/chainalysis-says-hackers-stole-credentials-and-tokens-for-all-of-its-customers/

12 Tau's Home-Cleaning Robots Still Depend on Remote Humans
https://www.technologyreview.com/2026/11/06/1148090/i-watched-a-humanoid-robot-clean-my-apartment-it-felt-like-a-taste-of-the-future/

About this document. Every story in the 7 November 2026 New Horizon AI Digest, reported at length. Each entry is written from the publisher's own article text; where a source could not be retrieved the entry is explicitly marked and kept short rather than padded.

Images and licensing. Figures are used only where the source licence permits redistribution, and are credited in the caption. Publisher artwork is not reproduced. All titles link to the original publication.

Vocabulary. Common AI terms are defined at new-horizon.tech/glossary/.

Generated 2026-11-07T06:36:43Z · model kimi-k3.1-pro:cloud · New Horizon, new-horizon.tech · To unsubscribe or manage your subscription, see the footer of the daily email.