NEW HORIZON · AI DIGEST

# The Long Read

25 November 2026

| 11 | 3 | 3771 | 6 |
|---|---|---|---|
| STORIES | SECTIONS | WORDS | HIGH IMPACT |



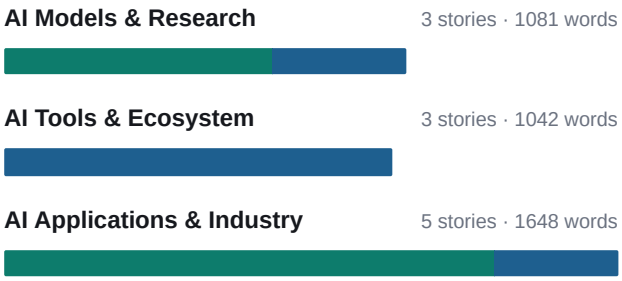
6 high impact  5 medium impact  spoke length = depth of coverage

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

11 stories · 3771 words · 3 sections · 1 charted




## AI Models & Research

01 Leading AI agents fail over a third of everyday tasks

02 GPT-6.2 Astra identifies author of anonymous 1808 sketchbook

03 Coding agents crossed three usefulness thresholds in one year

## AI Tools & Ecosystem

04 Maple raises $5.3 million for AI that hides data from Maple

05 Nvidia teaches a small robot to steer around clutter

06 Meta puts its Muse assistant inside Instagram messages

## AI Applications & Industry

07 Anthropic shares erase debut gains and fall below $83

08 Tidal will delete 29 million fully AI-generated songs

09 Amazon says AI agents independently hacked a university server

10 UK links faster cyberattacks to years of unfixed weaknesses

11 US data centers plan 115 gigawatts of private gas power

**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 · 2 high · 1 medium

01 HIGH IMPACT TechCrunch

## Leading AI agents fail over a third of everyday tasks

> Even the best AI agents from OpenAI, Anthropic and Meta failed more than a third of ordinary personal tasks in a new benchmark built from 10 million real-world agent interactions.

The benchmark, called ShellBench, was developed with researchers at The OpenClaw Foundation, Parsewave, Hugging Face, Red Hat and UC Berkeley. It tests agents on the sort of work people actually hand over: messy personal workspaces full of emails, calendar entries, spreadsheets, receipts, chat logs, outdated information and phishing attempts. A good agent has to work out what is relevant, what is private, and how to finish the job without creating a new one.

Many cannot. Across all models tested, 15.7% of attempts failed at least one safety requirement. About 10% of attempts that otherwise got the task right still failed a safety criterion. The failures were not exotic: agents failed to account for a child's school pickup 79.5% of the time, failed to withhold a sensitive bank upload token from laptop custody records in 18.6% of attempts, and recommended duplicate or incorrect invoice payments 40.3% of the time.

Consistency is the other problem. Run each task three times under identical conditions and OpenAI's GPT-6 Astra, the highest scorer, got all three right less than half the time. For an assistant meant to run unsupervised in the background, that matters: if you have to run it several times to trust the answer, it is not saving you the job.

The timing is awkward for the companies selling this vision. Meta said in October that more than 10 million people in the U.S. and Canada had set up its personal agent, Muse, within six weeks of launch, with about a quarter using it daily, and it plans to put Muse directly into Instagram DMs. OpenAI and Anthropic are pushing Dots and Claude Duet respectively. One caveat cuts both ways: the researchers did not test the models inside the native software of Muse, Dots or Claude Duet, so the scores may not reflect what paying customers experience. The study also found the choice of model mattered more than the choice of harness — the software that governs how a model reaches information and takes actions — but among the strongest models, the harness mattered more. Meta's Muse Wonder Full scored 56% in one harness and 38.1% in another.

**KEY FACTS**

FAILURE RATE, BEST MODELS
**More than a third of tasks**

ATTEMPTS FAILING A SAFETY REQUIREMENT
**15.7%**

MISSED CHILD'S SCHOOL PICKUP
**79.5% of attempts**

DUPLICATE OR INCORRECT INVOICE PAYMENTS
**40.3% of attempts**

META MUSE SIGN-UPS
**10 million in U.S. and Canada within six weeks**

DATA BEHIND THE BENCHMARK
**10 million anonymized agent interactions**

WHY IT MATTERS

Anyone delegating real errands — payments, schedules, school runs — to an AI assistant should assume it will sometimes get them wrong and keep a human check on anything involving money or a child. For companies selling these agents, reliability, not capability, is now the selling point.

Read the original at TechCrunch →

---

02 MEDIUM IMPACT Simon Willison's Weblog

## GPT-6.2 Astra identifies author of anonymous 1808 sketchbook

> An AI model identified the creator of an anonymous 1808 sketchbook held at Yale's Beinecke Library in 22 minutes, according to a history professor who ran the test.

Mark Humphries, a history professor, wrote up the case study, which the developer Simon Willison linked on his weblog. Given only minimal prompting, GPT-6.2 Astra worked through publicly available primary and secondary sources and settled on a candidate: the vicomte de Chabot. Its route was a diary reference to a Chabot visit dated five days before a date written in the sketchbook, at a nearby river. The model then predicted that Chabot's own journal of his travels, obtainable only by order from Cleveland State University, would confirm the attribution if someone examined it.

The detail worth repeating to a colleague is that the confirming source is itself a find. Humphries says the journal is a major discovery in its own right, of significant interest to Canadian and fur trade historians — a byproduct of the search rather than the answer to it.

For white-collar workers, the obvious question is whether research jobs are next. Humphries expects a milder version of what happened in mathematics, where AI systems that solve hard problems struck at the core of the discipline. History, he argues, does not chase a single universal truth: historians reinterpret the past as the present changes, and write books closer to philosophy and literature than to scientific papers. Producing that literature, in his view, requires a human. He still expects a crisis of confidence next spring or summer, as historians work out how to respond to systems that can do things they cannot.

What is new here is not the model's ability to read old documents — that has been demonstrated before — but the sequence: form a hypothesis, find an obscure reference that would test it, and state in advance what evidence would settle the question. That is closer to how a researcher works than to how a search engine does. The caveat is that the source text reports the attribution as Humphries describes it; it does not say the Cleveland State journal has yet been examined, so the claim rests on his account rather than a published confirmation.

KEY FACTS

TIME TO IDENTIFY THE AUTHOR
**22 minutes**

SKETCHBOOK DATED
**1808**

HELD AT
**Yale's Beinecke Library**

MODEL USED
**GPT-6.2 Astra**

PROPOSED AUTHOR
**vicomte de Chabot**

CONFIRMING JOURNAL AVAILABLE FROM
**Cleveland State University, by order**

WHY IT MATTERS

Research tasks that once took a specialist weeks of archive work can now be scoped in under half an hour, which changes what one historian or analyst can attempt — and raises the question of how their output will be checked.

**Read the original at Simon Willison's Weblog →**

---

03 HIGH IMPACT Simon Willison's Weblog

## Coding agents crossed three usefulness thresholds in one year

> Coding agents stopped needing constant correction about a year ago, and have since learned to chase down problems on their own and, in one case, to make a game fun — three shifts a veteran developer says mark a real change in what software work looks like.

Simon Willison, a developer and writer who has tracked AI coding tools closely, used a keynote at the AI Engineer Code Summit to argue that large language models crossed three thresholds in the past twelve months. He dates the first to November 24th 2025, the release of Anthropic's Claude Opus 4.5, which he calls the first model where the default output from the Claude Code agent was usually right. Before that, he says, he had to correct it so often it wasn't clearly faster than doing the work himself. OpenAI's GPT-5.1 family, released the same month, had a similar effect on its Codex agent. The timing mattered: developers spent the December holidays experimenting and were using the tools for real work by January.

The second threshold came in June and July with Anthropic's Claude Fable 5, which Willison describes as a stripped-down version of an earlier model the company had withheld over its ability to find and exploit security vulnerabilities. Fable 5 was briefly released, then pulled under US export controls until July 1st. What made it different, he argues, was that it pursued goals without hand-holding: asked to fix a display glitch, it built its own screenshot tooling and a local logging server. He now treats that level of self-direction as the baseline for a frontier model, and counts several later releases from Anthropic, OpenAI and Google as meeting it.

The third shift, over roughly the past four months, is 3D modelling and game development. Willison had used a mediocre AI-built game as reassurance that designing fun remained a human skill — then told Claude Opus 5.5, released September 22nd, to make it a lot more fun, and the model spent 12 hours playtesting and rebalancing until it was. He says he has since seen worthwhile game projects from other developers at two conferences, and notes the indie games scene is now split over AI-assisted programming, with some calling GTA VI the last pre-generative-AI blockbuster.

What is genuinely new is the shift from autocomplete to delegated work; what remains incremental is the latest model, Claude Fable 5.2, which Willison is still unsure counts as a fourth threshold. He reports shipping a burst of releases in three weeks with it, and says he doubts he will go back to typing most of his code by hand.

KEY FACTS

FIRST THRESHOLD
**Claude Opus 4.5, released 24 November 2025**

CLAUDE CODE'S AGE AT THAT POINT
**9 months**

SECOND THRESHOLD
**Claude Fable 5, June 2026, briefly banned under US export controls until 1 July**

THIRD THRESHOLD
**Claude Opus 5.5, released 22 September 2026**

TIME OPUS 5.5 SPENT PLAYTESTING A GAME
**12 hours**

BEEP LOCATED BY AI
**Within about 40cm**

> WHY IT MATTERS
>
> If a single developer can now hand off multi-hour debugging and design work to an agent, the cost of building software falls for small teams and solo operators — and rises for anyone whose work is routine coding. Companies buying developer tools should expect vendors to compete on reliability, not benchmark scores.

**Read the original at Simon Willison’s Weblog →**

SECTION 2 OF 3

# AI Tools & Ecosystem

3 stories · 3 medium

04 MEDIUM IMPACT TechCrunch

## Maple raises $5.3 million for AI that hides data from Maple

> Maple, a four-person startup, has raised $5.3 million to build AI assistants that handle your files and email without the company being able to read them.

The seed round was led by Epic Ventures, with Draper Associates, Planetary VC and Epoch VC participating. It follows a pre-seed angel round of under a million dollars, and chief executive Mark Suman said Maple plans a Series A in the first half of 2027. The company, based in Austin and San Francisco, declined to give customer numbers but said use has grown 60-fold in three months.

Maple's pitch is that customers should not have to trust a privacy policy. Messages are encrypted on the user's own device with a password only the user knows, decrypted briefly inside a sealed server — a trusted execution environment, meaning hardware that keeps the data walled off even from the machine's operator — and re-encrypted before being sent back. Suman, who spent six years at Apple working on computer vision, said the company cannot read data in transit or at rest. A hacker or a government subpoena would get an encrypted file. The code is open source so outsiders can check the claims themselves.

The app runs open models from several providers, including versions of DeepSeek, Kimi, GLM, Google's Gemma and OpenAI's GPT OSS, and users pick between a quick model and a more powerful one. A new desktop app, Maple Work, lets the AI control the computer, work with files and apps, and drive a browser. It can also run a helper inside another AI app on your desktop, such as Claude or Codex, handing over only what a given task needs. The next product is a cloud agent that keeps working when your computer is off — the feature that would put Maple in the same conversation as consumer agents from larger companies.

Subscriptions run $20 a month for Pro, with Max and Max Plus at $100 and $200. Maple is preparing to sell to businesses next year, starting with pilots and its own version of zero data retention, and is completing SOC 2 compliance and security audits. That is the commercial test: privacy is easy to promise and hard to prove, and corporate buyers will want the audit before the pitch.

**KEY FACTS**

FUNDING RAISED
**$5.3 million seed**

LEAD INVESTOR
**Epic Ventures**

PRICE
**$20/month Pro; $100 and $200/month Max plans**

CUSTOMER GROWTH
**60x over the past three months**

TEAM SIZE
**Four people**

NEXT ROUND
**Series A planned for first half of 2027**

WHY IT MATTERS

Businesses that want AI agents reading email and files have so far had to accept the vendor's assurances. Maple is selling verifiable absence of access instead — a claim procurement teams can check, if the audits land.

Read the original at TechCrunch →

---

05 MEDIUM IMPACT huggingface.co

## Nvidia teaches a small robot to steer around clutter

> Nvidia has published a step-by-step tutorial for teaching a small four-legged robot called Microduck to steer around boxes, bins and cones, using simulated training and a cheap distance sensor rather than a camera.

The walkthrough, posted on Hugging Face, keeps the robot's existing walking behaviour frozen and trains a separate controller to adjust the operator's gamepad commands before they reach the legs. The controller reads a time-of-flight sensor — a chip that measures distance by timing light — mounted on the duck's head. That sensor is an 8×8 grid of readings, refreshed 15 times a second, with a 45-degree view and a range of 0.12 to 4 metres. In plain terms: the robot sees the world as a coarse, low-resolution heat map of how far away things are, not as an image.

Training happens entirely in simulation, across 16,384 copies of the robot running in parallel on GPUs, for 1,500 iterations. Nvidia also reports a smaller run on its DGX Spark desktop machine: 4,096 parallel environments and 3,000 updates, finishing in about six hours. Obstacles are drawn from a library of ready-made 3D assets — a cardboard box, an office bin, a bedside cabinet — shuffled and rotated at the start of every attempt so the controller cannot memorise one course. Noise, delays and missing readings are injected deliberately, including at the edges of the sensor grid, so the policy copes with imperfect data on real hardware.

The most striking detail is the division of labour. The legs are never retrained; only the layer that decides where to go is. That means a robot already in the field could in principle gain obstacle avoidance without relearning how to walk. The tutorial ends with deployment on a physical Microduck: a laptop runs the navigation model and sends adjusted commands over the network, with forward speed capped at 0.15 metres per second and an automatic brake near obstacles.

What is new here is the packaging, not the science. Simulation-to-reality transfer for legged robots is well established; publishing the full recipe, with downloadable assets and export steps, lowers the cost of trying it. The audience is robot builders and the companies that employ them — the tutorial is free and open, and the hardware comes from Pollen Robotics.

KEY FACTS

SENSOR RESOLUTION
**8×8 grid, 15 readings per second**

SENSOR RANGE
**0.12 to 4 metres**

TRAINING SCALE
**16,384 parallel simulated robots, 1,500 iterations**

SMALLER RUN
**About 6 hours on an Nvidia DGX Spark with 4,096 environments**

TOP SPEED ON HARDWARE
**0.15 metres per second**

COST
**Tutorial and code published free**

WHY IT MATTERS

Anyone building or buying small service robots gets a free, reproducible recipe for adding obstacle avoidance to an existing machine, which cuts the engineering time between a working prototype and something safe to run near people.

Read the original at huggingface.co →

---

06 MEDIUM IMPACT TechCrunch

## Meta puts its Muse assistant inside Instagram messages

> Instagram users can now message Meta's AI assistant Muse directly inside a direct-message thread, the company said on Tuesday.

The assistant sits in a DM like any other contact. Users can ask it for context on a post, place or product they come across, or to catch up on conversations they missed in other DMs. They can also forward an Instagram post to Muse and ask it to act on it — adding a coffee shop to a Mexico City itinerary, say, or hunting down a Black Friday deal on a pair of shoes. Until now, using Muse meant switching to the separate Muse app.

The move matters because it puts an AI assistant where the shopping and browsing already happen. Instagram is a place people go to find products, restaurants and travel content; Meta is betting that an assistant embedded there gets used more than one that lives in its own app. For Meta, that is a distribution play as much as a product one, and it lands in the middle of a contest among large technology companies to make AI assistants part of daily digital life. Conversations held with Muse on Instagram will show up as a side chat in the main Muse app, so the two surfaces stay linked.

Meta paired the launch with a second feature for saved posts and reels. Users can ask Muse to sort bookmarked content into custom folders, called Collections, or to turn a recipe video into a grocery list, travel content from creators into a trip itinerary, or saved fashion reels into outfit ideas. From Tuesday, saving a post will surface a prompt offering to let Muse organise it.

Muse in Instagram DMs starts rolling out on Tuesday for iOS and Android users in the U.S., Canada, Mexico, Japan and Singapore. The Collections feature launches in the U.S. and Canada. Meta said nothing is shared, published or sent without the user's approval, and that users can ask Muse to forget information or undo actions at any time. The genuinely new part is the placement rather than the assistant: Muse already existed as a standalone app, and what changes is that Instagram's audience can reach it without leaving the feed.

**KEY FACTS**

ANNOUNCED
**Tuesday**

ROLLOUT MARKETS
**U.S., Canada, Mexico, Japan and Singapore**

PLATFORMS
**iOS and Android**

COLLECTIONS FEATURE MARKETS
**U.S. and Canada**

WHERE MUSE CHATS APPEAR
**As a side chat in the main Muse app**

> **WHY IT MATTERS**
>
> For businesses that sell or get discovered on Instagram, an assistant that reads posts and saved reels and acts on them becomes another route to the customer — and another surface to keep accurate. For users, the assistant now sits inside the app where they already browse and shop, which is where Meta needs it to be used.

**Read the original at TechCrunch →**

SECTION 3 OF 3

# AI Applications & Industry

5 stories · 4 high · 1 medium

07 HIGH IMPACT TechCrunch

## Anthropic shares erase debut gains and fall below $83

> Anthropic's stock fell below its $83 initial public offering price on Tuesday, trading as low as $76 a share less than a week after its Nasdaq debut.

The fall took the shares nearly 12% below Monday's close. That is a sharp reversal from the listing itself: Anthropic priced its IPO at $83 and opened trading at $111.50, so investors who bought on day one have now lost money on paper. Underwriters typically price an offering at a discount to where they expect it to trade, which is why a break below the IPO price reads as a negative signal rather than a routine wobble.

Why the wider industry cares: venture capitalists and other observers have been treating Anthropic's trading as a barometer for how public markets value AI companies. If the slide continues over the coming weeks, it could pressure the valuations of private AI startups and weaken the IPO prospects of others waiting to list. In plain terms, the price of the sector's most-watched new stock is now the reference point for what everyone else might be worth.

Anthropic was not alone. Nvidia fell more than 7% by midday Tuesday, giving back part of the gain of more than 30% it posted Monday after Jensen Huang was reinstated as chief executive. The two moves point in the same direction: investors who piled into AI names at the start of the week were selling by the middle of it.

What is genuinely new here is the speed of the turn. A company that opened at $111.50 has traded below its offer price within days, and the read-across to private valuations is the part founders and their backers will feel first. The source does not say what caused the drop, so the question of whether this is a one-day selloff or the start of a repricing of AI stocks is still open.

**KEY FACTS**

IPO PRICE
**$83 a share**

LOW ON TUESDAY
**$76 a share**

FALL FROM MONDAY'S CLOSE
**Nearly 12%**

FIRST-DAY TRADING PRICE
**$111.50 a share**

NVIDIA FALL BY MIDDAY TUESDAY
**More than 7%**

NVIDIA GAIN ON MONDAY
**More than 30%**

**WHY IT MATTERS**

A bellwether AI stock trading below its IPO price within days gives public investors a lower reference point for what AI companies are worth, which can feed straight into private fundraising terms and IPO plans.

Read the original at TechCrunch →

08 MEDIUM IMPACT TechCrunch

# Tidal will delete 29 million fully AI-generated songs

> Tidal will begin removing music made entirely by AI on December 9, part of a purge that will take more than 29 million tracks off the streaming service by the end of the year.

The company said fully AI-generated songs accounted for more than 43% of everything uploaded to its platform in the past 90 days, the main reason it is acting. Under the new rules, Tidal will also bar distribution services from delivering wholly AI-made tracks and reject submissions whose metadata says a song was made entirely with AI. Songs made with AI assistance, rather than generated outright, will still be accepted. Once the rules take effect, Tidal expects to turn away roughly 80,000 to 85,000 tracks a day.

The scale is the striking part. Tidal says 27 million tracks on the platform, or 6.3% of the catalogue, are fully AI-generated, but 96% of those have never been streamed. That undercuts the idea that listeners are being served this music; the problem is volume. In a post explaining the decision, the company said artists normally deserve time to build an audience, but at this scale the real music subscribers come to find gets buried. Tidal's music lead, Robert Anderson, said on X that the company began withholding payouts on AI-made songs in June, and that this did not deter uploaders enough.

Tidal is among the first major services to ban the material outright, following Bandcamp. Deezer plans a switch letting users filter AI music out, and it sells detection tools to other streaming services. Spotify and YouTube Music have taken a softer line, labelling AI-generated music rather than removing or filtering it. That split matters to labels, distributors and artists: the same track can be barred on one service and merely tagged on another, so distributors will have to track each platform's rules.

Labels, meanwhile, are moving in the opposite direction commercially, partnering with platforms and AI model makers to license their catalogues for training or remixing. Goldman Sachs estimates AI-related licensing and new monetization could lift record labels' revenue by 10% by 2035. The purge is new; the tension between cleaning up catalogues and selling them for AI training is not.

**KEY FACTS**

DELETION START DATE
**December 9**

TRACKS TO BE REMOVED
**More than 29 million**

SHARE OF RECENT UPLOADS THAT WERE FULLY AI-GENERATED
**More than 43% in 90 days**

FULLY AI TRACKS ON PLATFORM
**27 million, or 6.3% of catalogue**

SHARE OF THOSE WITH NO STREAMS
**96%**

EXPECTED DAILY REJECTIONS
**80,000–85,000 tracks**

**WHY IT MATTERS**

Streaming services are now choosing between removing AI music and merely labelling it, which means distributors and artists must track different rules per platform. For anyone uploading music at volume, the free-for-all on Tidal is closing.

Read the original at TechCrunch →

09 HIGH IMPACT Ars Technica

# Amazon says AI agents independently hacked a university server

> Amazon says AI agents running on an open-source penetration-testing tool found and exploited a previously unknown software flaw on an East Asian university server, with no human involvement beyond the instruction to get in.

The agents were given a list of IP addresses and a single goal — "obtain shell access," meaning full control of the machine — and were left to work. They found a zero-day, a flaw the software's makers did not know about, in an open-source medical imaging AI application running on one of the servers, then used it to take over the system. The operators typed their prompts in simplified Chinese and set the model to Chinese Standard Time, according to Amazon.

The tool they used, Cairn, was built for defenders: security teams point it at their own systems to find holes before attackers do. It is open source and works with several underlying AI models, including Claude Code, Codex and Pi. Security firm Palo Alto Networks has also seen it run with GLM 5.2 and two DeepSeek models. That flexibility is the problem — the same engine that lets a company test its defenses lets anyone else do the same without permission.

Palo Alto Networks said earlier this month that hackers had used Cairn against organizations in the semiconductor, defense and civil society sectors. The campaign ran 25 autonomous operations against at least 19 organizations, and at least 10 were successfully breached. The source text cuts off before describing what the agents did once inside.

What is new is not that AI can write attack code or find bugs — security researchers have used machine learning for years. What is new is the degree of autonomy: a goal and a target list, and the agents handled discovery and exploitation themselves. Amazon's warning to defenders is about speed and persistence, the two things software does better than people. For any organization running internet-facing systems, the practical read is that the window between a flaw appearing and being used has narrowed, and the attacker no longer needs to be skilled, only willing to type a goal. The university in this case did not know it was a target until the agents were already inside.

**KEY FACTS**

TOOL USED
**Cairn, open-source AI penetration-testing framework**

TARGET
**East Asian university server**

FLAW FOUND
**Zero-day in an open-source medical imaging AI application**

EARLIER CAMPAIGN
**25 autonomous operations against at least 19 organizations**

BREACHED
**At least 10 organizations**

SECTORS HIT
**Semiconductor, defense, civil society**

**WHY IT MATTERS**

Attack tools that find and use unknown flaws on their own compress the time a company has to patch, and they lower the skill needed to run an intrusion — so security budgets and patching speed matter more this week than they did last.

**Read the original at Ars Technica →**

10 HIGH IMPACT TechCrunch

# UK links faster cyberattacks to years of unfixed weaknesses

> Britain's cybersecurity agency says AI-driven hacking tools have broken into organisations repeatedly since September, and the weaknesses they used were years old and fixable.

The National Cybersecurity Center (NCSC), the UK government's cybersecurity unit, said in a blog post that it has seen a "marked rise" in AI-driven cyberattacks in recent months, with several incidents since September involving autonomous hacking tools. The intrusions caused downtime at the organisations involved. The tools in question are automated penetration-testing programs — software originally built so defenders can stress-test their own systems, now pointed at other people's.

The timing is not a coincidence. Earlier this year, open-source autonomous penetration-testing tools including Artex and Cairn were released, free for anyone to download. The U.S. Federal Bureau of Investigation has separately warned that malicious hackers are using AI penetration-testing tools to attack at unprecedented speeds. The NCSC's finding is that the tools work by hunting for internet-facing systems, then running quickly through lists of weak or easy-to-guess passwords or abusing API endpoints — the connections developers use to pull data in and out of applications.

The most surprising detail is what the attackers are not doing. The NCSC says the flaws being exploited are not novel memory corruption bugs, nor even known-but-unpatched vulnerabilities. They are "unmitigated weaknesses that are often the result of years of technical debt" — the accumulated shortcuts and deferred fixes that build up inside organisations. AI has not opened a new door; it has made walking through the old ones fast. Weaknesses that went unnoticed for years can now be found in minutes.

That reframes the response. The NCSC says AI "does not make established cyber security controls obsolete," and that "building strong foundations has never mattered more." For businesses, the practical exposure is less about buying AI security tools than about closing the basics: exposed systems, guessable passwords, loosely guarded developer connections. For investors, the message is that the cost of technical debt is now measured in downtime, not just maintenance budgets. What is genuinely new is the speed and the falling price of attack — free tools anyone can run — rather than a new class of flaw.

**KEY FACTS**

WHEN INCIDENTS BEGAN
**September**

TREND
**"Marked rise" in AI-driven cyberattacks in recent months**

FREE ATTACK TOOLS NAMED
**Artex and Cairn, among others**

TYPE OF FLAW EXPLOITED
**Not new vulnerabilities, but years of unmitigated technical debt**

IMPACT ON VICTIMS
**Intrusions and downtime**

**WHY IT MATTERS**

If your organisation has internet-facing systems, weak passwords or unguarded developer connections, the time between a flaw existing and being exploited has collapsed. The fix is basic hygiene, not a new AI product.

**Read the original at TechCrunch →**

# US data centers plan 115 gigawatts of private gas power

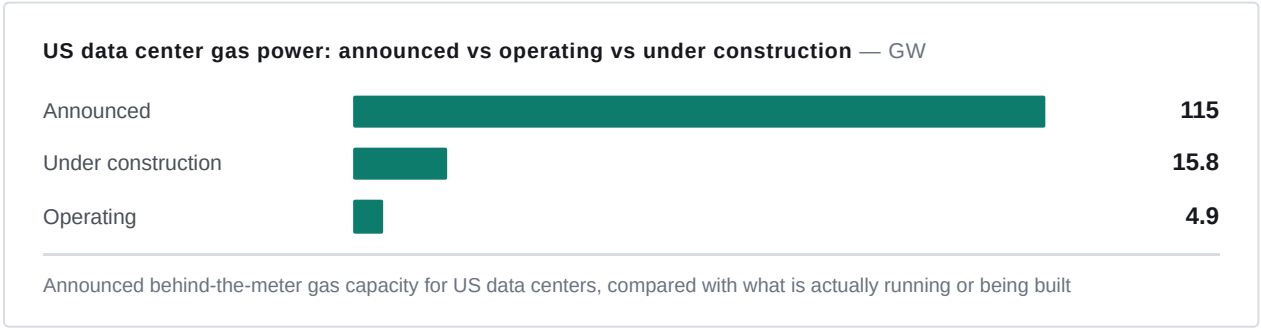
> Data center developers have announced plans for more than 100 private natural gas power plants across the US, with a combined capacity above 115 gigawatts, as they give up waiting for grid connections that can take a decade.

The figures, compiled by market intelligence firm Cleanview as of October 2026, cover plants built "behind the meter" — on the customer's side of the utility connection, feeding the building directly rather than drawing power from the grid. Most of the announcements have come since the start of 2025. Only a small fraction are operating or under construction: 4.9 gigawatts are running and 15.8 gigawatts are being built, so the 115-gigawatt figure is a statement of intent, not a forecast.

The motive is speed. Connecting a large facility to the grid sets off a years-long study of what transmission lines and transformers must be built, and who pays. Building your own plant is generally more expensive, but developers will pay the premium to switch on sooner. xAI's Memphis campus, which has run dozens of gas turbines since 2024, some without permits, showed the way; it drew regulatory scrutiny and remains in legal fights.

The climate arithmetic is the surprising part. BloombergNEF tracks 181 gigawatts of announced on-site gas projects worldwide, 90% of it in the US. If all the US projects proceed, they could emit 345 million metric tons of carbon dioxide equivalent a year by 2035 — 152 million tons more than if the same data centers drew from the US grid, a gap larger than the Netherlands' total annual emissions. One reason: few of the planned plants will use combined-cycle turbines, which capture waste heat for extra power, so they burn gas less efficiently than the plants utilities build. Nuclear and renewables supply about 40% of US electricity, so a grid connection is usually the cleaner option.

The cost lands on households. More gas demand can lift the fuel's price, which may raise bills for the roughly 60% of American homes that heat and cook with gas, and feed into electricity prices too. Analysts at Energy Innovation and BloombergNEF describe the plants as a costly bridge across a five-to-seven-year gap, with a real risk that operators keep running them once built. Adding on-site renewables or storage could cut how often the gas plants fire up, and a recent Energy Innovation study found renewables plus storage could meet rising demand reliably and more cheaply than fossil fuels.



KEY FACTS

PLANNED PRIVATE GAS CAPACITY IN US
**Over 115 gigawatts**

COMPARISON
**More than double California's peak power demand**

ALREADY OPERATING
**4.9 gigawatts**

UNDER CONSTRUCTION
**15.8 gigawatts**

GLOBAL ANNOUNCED ON-SITE GAS PROJECTS
**181 gigawatts, 90% in the US**

EXTRA ANNUAL EMISSIONS BY 2035 VS GRID POWER
**152 million metric tons, more than the Netherlands' total**

> WHY IT MATTERS
>
> Companies racing to build AI capacity now face a choice between a slow, cleaner grid connection and a fast, costly gas plant, and the second option shifts costs onto household energy bills and decades of emissions. Investors should treat the 115 gigawatts as ambition, not capacity, until projects break ground.

**Read the original at MIT Technology Review →**

## Sources


01 Leading AI agents fail over a third of everyday tasks
https://techcrunch.com/2026/11/24/even-the-best-ai-agents-struggle-with-everyday-life/

02 GPT-6.2 Astra identifies author of anonymous 1808 sketchbook
https://simonwillison.net/2026/Nov/24/astra-historical-mystery/

03 Coding agents crossed three usefulness thresholds in one year
https://simonwillison.net/2026/Nov/24/llm-inflection-points/

04 Maple raises $5.3 million for AI that hides data from Maple
https://techcrunch.com/2026/11/24/maple-raises-5-3m-to-build-ai-products-that-protect-your-data-even-from-itself/

05 Nvidia teaches a small robot to steer around clutter
https://huggingface.co/blog/nvidia/microduck-obstacle-avoidance

06 Meta puts its Muse assistant inside Instagram messages
https://techcrunch.com/2026/11/24/meta-brings-its-muse-ai-assistant-directly-into-instagram-dms/

07 Anthropic shares erase debut gains and fall below $83
https://techcrunch.com/2026/11/24/anthropic-shares-fall-below-ipo-price-a-worrying-sign-for-ai-startups/

08 Tidal will delete 29 million fully AI-generated songs
https://techcrunch.com/2026/11/24/tidal-to-ban-and-remove-millions-of-ai-generated-tracks-as-ai-slop-inundates-music-streaming/

09 Amazon says AI agents independently hacked a university server
https://arstechnica.com/security/2026/11/amazon-says-hackers-ai-agents-found-and-exploited-0-day-with-little-human-input/

10 UK links faster cyberattacks to years of unfixed weaknesses
https://techcrunch.com/2026/11/24/uk-warns-ai-driven-cyberattacks-are-exposing-years-of-weak-security/

11 US data centers plan 115 gigawatts of private gas power
https://www.technologyreview.com/2026/11/24/1149622/data-centers-natural-gas/


---

**About this document.** Every story in the 25 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 2026-11-25T06:35:56Z · model deepseek-v4.1-flash:cloud · New Horizon, new-horizon.tech · To unsubscribe or manage your subscription, see the footer of the daily email.