

The AG Gazette

Models, machines, and matters of consequence

PARIS, MONDAY 19 OCTOBER 2026

AFTERNOON EDITION · No. 2026-10-19-AFTERNOON

ISSUED 16:09 CEST

PRICE: COMPLIMENTARY

AFTERNOON EDITION — MONDAY 19 OCTOBER 2026

Reporting window: 19 Oct 2026, 06:08 CEST to 19 Oct 2026, 16:09 CEST (Paris time). Issued 19 Oct 2026, 16:09 CEST.

This edition follows the Monday morning issue, published at 06:08 CEST. Five verified developments were published on Monday 19 October after that issue: MiniMax's M3.1 Flash model, Amazon's TorchNeuron backend for running PyTorch on Trainium, NorForge's investigation into exposed AI-agent attack panels, a Homeland Security order removing DeepSeek products from civilian federal IT, and Tenstorrent's \$1.5 billion funding round. No older context is needed, so there is no BACKGROUND section. Stories from earlier editions are not repeated.

5 NEW OR UPDATED DISPATCHES · 0 BACKGROUND ITEMS · PREVIOUS: [2026-10-19-MORNING](#)

THE NEWS

MINIMAX RELEASES M3.1 FLASH AND CONFIRMS THE ANONYMOUS 'SPACE BUNNY' MODEL WAS AN EARLY VERSION

PUBLISHED: 19 Oct 2026

MiniMax said on 19 October that it is rolling out M3.1 Flash through the MiniMax API, its M Plan subscription and its MiniMax Code app. The company said Space Bunny, a model that processed more than 200 trillion tokens while operating anonymously, was an early version of M3.1 Flash. MiniMax said model weights will be released later, and that a larger Pro model will also arrive later.

MiniMax lists flat pricing of \$0.15 per million input tokens, \$0.60 per million output tokens and \$0.015 per million cached input tokens. It claims online inference is nearly three times faster and agentic-task usage costs are 70% lower than its predecessor, with output speeds above 200 tokens per second and a median above 170.

In its own benchmark results, MiniMax reports 73.7% on DeepSWE v1.1, 61.7 on OSWorld V2, 29.6 on Agents' Last Exam, 76.2 on BankerToolBench and 82% on CyberGym.

MarkTechPost reports that MiniMax's product page lists 428 billion total parameters with about 23 billion active, a one-million-token context window and reasoning that cannot be switched off. It says the weights are due on 23 October under a license not yet disclosed, and that MiniMax's comparison table puts the model at 38.1 on Terminal Bench 4.0.

ANALYSIS — WHY IT MATTERS. *Analysis: Chinese labs keep cutting the price of agent-oriented models, and testing them under anonymous names before launch makes it harder for developers to know which build they actually evaluated.*

CAVEAT. All benchmark numbers are vendor-reported; this edition found no independent replication. The 23 October weights date comes from MarkTechPost, while MiniMax's announcement says only 'later'.

SOURCES

- MiniMax** <<https://www.minimax.io/blog/minimax-m31f>> (primary source) **CHECKED: 19 Oct 2026, 16:05 CEST**
- MarkTechPost** <<https://www.marktechpost.com/2026/10/19/minimax-m3-1-flash-release/>> (reporting) **CHECKED: 19 Oct 2026, 16:05 CEST**

AMAZON SAYS TORCHNEURON LETS EXISTING PYTORCH CODE RUN ON ITS TRAINIUM CHIPS, WITH AN OPEN-SOURCE RELEASE SET FOR 20 OCTOBER

PUBLISHED: 19 Oct 2026, 15:00 CEST

Amazon announced TorchNeuron, a native PyTorch backend for its AWS Trainium AI chips. Amazon says PyTorch code that developers already write runs on Trainium with a single configuration change and no rewrite, and that in a side-by-side demonstration a training script learned at the same rate on a conventional chip and on Trainium.

A post on the PyTorch Foundation blog by AWS engineers says TorchNeuron plugs into PyTorch's PrivateUse1 device backend and torch.accelerator APIs. It supports eager execution, torch.compile and the Neuron Kernel Interface for custom kernels on Trainium2 and Trainium3. The post says the code will be released under the Apache 2.0 license at aws-neuron/torch-neuronx as a public beta on 20 October, at PyTorch Conference North America.

Amazon says it is contributing changes upstream to make PyTorch less tied to specific chips and tests TorchNeuron against hundreds of thousands of models on every build. It names Cartesia and the Lieber Institute for Brain Development among early users.

ANALYSIS — WHY IT MATTERS. *Analysis: much of Nvidia's lead rests on software that developers would have to rewrite to leave; lowering that porting cost is how rival chipmakers compete.*

CAVEAT. The no-rewrite and training-parity claims are Amazon's own and were not independently tested. The open-source release was scheduled for the day after this edition and could not be checked.

SOURCES

- Amazon** <<https://www.aboutamazon.com/news/innovation-at-amazon/torchneuron-pytorch-aws-trainium-chips-ai>> (primary source) **CHECKED: 19 Oct 2026, 16:05 CEST**
- PyTorch Foundation blog** <<https://pytorch.org/blog/introducing-torch-neuron-native-pytorch-backend-for-aws-trainium/>> (primary source) **CHECKED: 19 Oct 2026, 16:05 CEST**

EXPOSED AI-AGENT ATTACK PANELS SHOW MORE THAN 1,100 INTRUSION PROJECTS, SECURITY FIRM NORFORGE REPORTS

PUBLISHED: 19 Oct 2026

NorForge published an investigation on 19 October into publicly exposed control panels for Cairn, an open-source AI agent platform it describes as a state-based problem-solving engine. An operator sets a starting point and a goal, and agent workers move toward it, posting discovered facts and planned next steps to a shared board.

NorForge says it found more than 20 exposed Cairn panels it assesses as likely malicious, all apparently financially motivated, with no obvious state-sponsored activity. Its largest cluster, which it labels AIA0001, uses Chinese, spans 14 panels and more than 1,100 attack projects, and mostly targets organizations in the United States, Japan and Australia, apparently seeking persistent access to e-commerce and payment infrastructure. Some exposed panels appeared to belong to legitimate red teams, which NorForge says could put their clients at risk.

The Verge, citing Reuters, reports that victims included John Deere, SharkNinja and Cummins, that the most active attacker broke into more than 100 websites, and that successful breaches took about 23 hours. John Deere told Reuters there was 'no indication that sensitive information was accessed', and SharkNinja said a standalone product-support site was affected and the threat contained. The Verge reports that Cairn's creator said earlier this month that the platform 'was never a hacking tool'.

ANALYSIS — WHY IT MATTERS. *Analysis: agent frameworks cut the human labor behind large intrusion campaigns; in this case the attackers' own unsecured dashboards were what exposed them.*

CAVEAT. This is a security vendor's threat-intelligence report, not peer-reviewed research, and NorForge redacted panel addresses. Victim names come from Reuters via The Verge; Reuters itself returned HTTP 401 and was not read directly.

SOURCES

- NorForge** <https://norforge.com/blog/inside_cairn> (primary source) **CHECKED: 19 Oct 2026, 16:06 CEST**
- The Verge (citing Reuters)** <<https://www.theverge.com/ai-artificial-intelligence/1013643/ai-agents-hack-john-deere-shark-ninja-cummins>> (reporting) **CHECKED: 19 Oct 2026, 16:05 CEST**

US HOMELAND SECURITY ORDERS DEEPSEEK PRODUCTS REMOVED FROM CIVILIAN FEDERAL IT AND BARRED FROM PROCUREMENT

PUBLISHED: 19 Oct 2026, 14:45 CEST

The US Department of Homeland Security filed a notice for public inspection on 19 October, due to be published in the Federal Register on 20 October. It says the Secretary of Homeland Security has ordered two things: information technology products and services developed by Hangzhou DeepSeek Artificial Intelligence Co., Ltd, including affiliates, are excluded from federal civilian agency procurements; and all DeepSeek and DeepSeek-affiliate products, and third-party products and services developed using DeepSeek implementations, are to be removed from civilian agency IT.

The notice cites the Federal Acquisition Supply Chain Security Act of 2018 and a review of a Federal Acquisition Security Council recommendation. The signed order is filed under docket DHS-2026-0628, and a CISA team is named as the implementation contact.

Reuters, as published by CNA, reports that DHS said DeepSeek models pose 'multiple threat vectors' and that their use 'may result in the transfer of sensitive data to the PRC if used by U.S. government employees or vendors.'

ANALYSIS — WHY IT MATTERS. *Analysis: the clause covering third-party products 'developed using DeepSeek implementations' could reach vendors that build on DeepSeek's openly released models, not only DeepSeek's own app.*

CAVEAT. This edition read the public-inspection notice, not the signed order in its appendix, so deadlines and the exact scope of 'developed using DeepSeek implementations' were not checked. The order covers federal civilian executive-branch agencies; it is not a ban on private use.

SOURCES

- Federal Register (DHS notice 2026-21447)** <<https://public-inspection.federalregister.gov/2026-21447.pdf>> (primary source) **CHECKED: 19 Oct 2026, 16:07 CEST**
- CNA (Reuters)** <<https://www.channelnewsasia.com/business/us-banning-chinese-ai-deepseek-government-agency-it-systems-6472301>> (reporting) **CHECKED: 19 Oct 2026, 16:05 CEST**

TENSTORRENT RAISES \$1.5 BILLION AT A \$6.2 BILLION VALUATION TO SCALE ITS AI SERVERS AND CHIP IP

PUBLISHED: 19 Oct 2026

Tenstorrent said on 19 October that it has closed a \$1.5 billion Series E round at a \$6.2 billion post-money valuation. Trace Capital and Innovation X led the round, with participants including ARK Invest, Fidelity Management and Research Company, Morgan Stanley, Sandisk, T. Rowe Price, Tether Investments and XTX Markets.

The company says it will use the money to deploy its Galaxy servers at scale, keep serving open-weight language and video-generation models, and develop the AI and CPU designs it licenses to chipmakers. It names Samsung and LG as licensees of its TT-Ascalon RISC-V CPU and Tensix AI cores. Chief executive Jim Keller said: 'We built one architecture and sell it at every size.'

ANALYSIS — WHY IT MATTERS. *Analysis: investors are still funding alternatives to Nvidia, and Tenstorrent's pitch rests on open-source software and licensable designs rather than on chips alone.*

CAVEAT. Funding, valuation and investor list come from the company's announcement and were not independently confirmed.

SOURCES

- Tenstorrent** <<https://tenstorrent.com/newsroom/tenstorrent-announces-series-e>> (primary source) **CHECKED: 19 Oct 2026, 16:06 CEST**

KEY PATTERNS — EDITORIAL ANALYSIS

- The cost of running AI keeps falling from both ends: MiniMax prices an agent-oriented model at \$0.15 per million input tokens, and Amazon is trying to make moving PyTorch code off Nvidia hardware nearly free.
- AI is increasingly treated as a security and supply-chain risk, from the Homeland Security order removing DeepSeek products to NorForge's evidence of agent-driven intrusion campaigns.
- Money is still flowing to Nvidia alternatives, with Tenstorrent's \$1.5 billion round and Amazon's Trainium software push on the same day.

SOURCE COVERAGE & COLOPHON

COLLECTION ENGINE

VERSION: 3.28.0
RAN THIS EDITION: yes
ENGINE LOOK-BACK: 2 days

SOURCE STATUS

REDDIT: ok
YOUTUBE: no-results
HACKERNEWS: ok
POLYMARKET: timeout
GITHUB: partial
DEVTO: timeout

COVERAGE NOTES

- The fresh window runs from the morning issue (06:08 CEST, 19 October) to this issue. Two dispatches carry date-only publication stamps from their primary sources; coverage timestamps from other outlets place both after the morning issue.
- The installed last30days engine completed with 58 evidence items from Reddit, Hacker News and GitHub, followed by three web supplements and direct checks of primary sources. Evidence-item counts are not verified-development counts. None of the returned community comments addressed the stories selected here, so no community quotes are printed.
- GitHub returned partial results because its issue and pull-request searches failed with network errors. dev.to timed out. Polymarket timed out during TLS handshakes and also failed certificate host-name checks; verification was not bypassed. YouTube completed with no results. X was not searched. Failed or missing coverage is not read as an absence of discussion.

ENGINE FOOTER

All agents reported back!
└─ Reddit: 23 threads | 583 upvotes | 518 comments
└─ HN: 11 stories | 2,420 points | 1,113 comments
└─ GitHub: 24 items | 6 reactions | 51 comments
└─ Top voices: r/artificial, r/OpenAI, r/singularity

- The configured web search backend failed and its keyless fallback served the supplements. Public HTTPS pages were fetched directly with normal certificate validation. Reuters returned HTTP 401 and Investing.com HTTP 403, and the Federal Register website showed a CAPTCHA, so its public API and the notice PDF were used instead. Reuters facts are taken from outlets that republish or cite it.
- Excluded after checks: Google's Gemini 4 agents completing Minecraft (the underlying paper and its date were not verified), Anthropic's acquisition of Opaque Systems (announced by Opaque the previous week), MITRE ATT&CK v20 (released the previous week), Microsoft's 'Securing the agentic insider' post (commentary), Siemens and Santander partnerships with Nvidia (partnership announcements), a reported eight-state lawsuit against OpenAI (no primary filing checked in time), and further reactions to stories already covered, such as Anthropic's Public Benefit Fund and DeepSeek V4.1 Flash.
- No browser-cookie access, optional service or credential signup, paid enrichment, TLS bypass, Job Alerts edits or schedule changes were used. Research logs and full provenance remain private.

THIS EDITION

- [Exact input JSON](#)
- [Markdown text](#)
- [Permanent page](#)

The AI Gazette is a locally rendered digest. Each dispatch separates reported facts, labeled analysis and caveats, and cites the sources checked at the times shown. Background items are older or undated context, not new reporting.

Rendered as static HTML with inline styles and no scripts, cookies or analytics. Typefaces are requested from Google Fonts (fonts.googleapis.com, fonts.gstatic.com); if that request is blocked, Georgia, Times or another installed serif is used.