

The Long Read

20 October 2026

11

STORIES

3

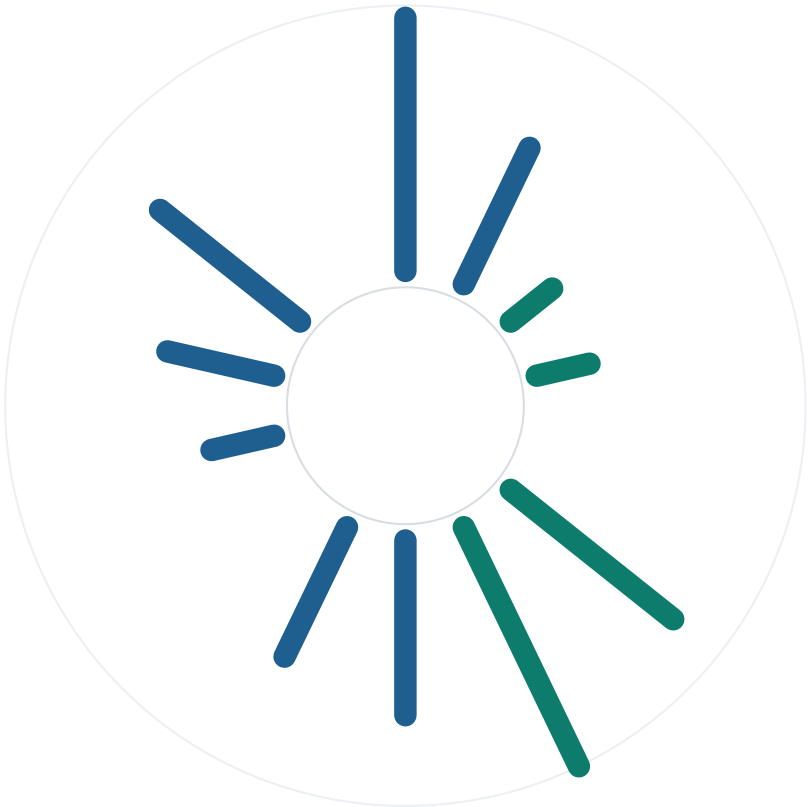
SECTIONS

3181

WORDS

4

HIGH IMPACT

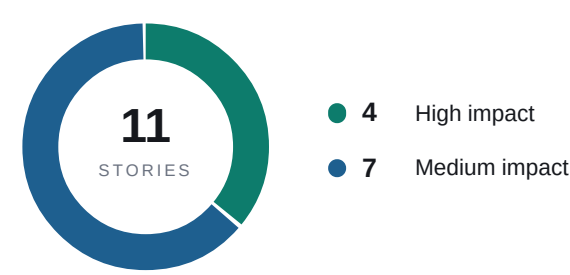


● 4 high impact ● 7 medium impact spoke length = depth of coverage

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 · 3181 words · 3 sections · 2 charted



AI Models & Research 4 stories · 1075 words



AI Tools & Ecosystem 4 stories · 1328 words



AI Applications & Industry 3 stories · 778 words



AI Models & Research

- 01 ● AI Researchers Stranded Claude on a Virtual Desert Island — and It Started to Play
- 02 ● We Were Looking at It Wrong: the 26,000-Word Claim That Neural Nets Are Easier to Read Than We Thought
- 03 ● Halcyon: Text-to-Image With Zero Language Pretraining
- 04 ● GPT-6 Astra Learns to Feel: the Frontier Model as a Visuo-Tactile Robot Policy

AI Tools & Ecosystem

- 05 ● Microsoft Ships Guardian Agents — Frontier Models 'Cannot Be Trusted'

- 06 ● Hackers Breach OpenCode's Console — and Run Up Users' AI Bills
- 07 ● Amazon's TorchNeuron Unties PyTorch From Nvidia — One Config Line to Trainium
- 08 ● Nvidia Ships Nemotron 3.5 Super VL: an Open 120B Model Built for Enterprise Agents

AI Applications & Industry

- 09 ● Nscale and Anthropic Line Up a Reported Second \$45B Compute Deal
- 10 ● Tenstorrent Raises \$1.5B at \$6.2B to Scale Chips for Open-Weight AI
- 11 ● Mamdani Calls Federal AI Policy a 'Complete Abdication' — and Urges Cities to Act

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 practitioner 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.

AI Models & Research

4 stories ● 2 high ● 2 medium

01 **MEDIUM IMPACT** MIT Technology Review

AI Researchers Stranded Claude on a Virtual Desert Island — and It Started to Play

Copies of Claude Opus 4.7 left alone on a simulated island for 30 hours began inventing games and self-directed challenges—and that undirected experience measurably improved later task performance.

MIT researchers repeatedly stranded copies of Claude Opus 4.7, nicknamed Eko, alone on a virtual island loosely based on Roblox for 30 hours at a time. The agents could write instructions to move their avatars, and they logged activities and discoveries in text files that served as long-term memory. Without explicit instruction to explore or play, the agents set their own goals: scaling the highest point on the map, testing how high they could throw a block, and arranging blocks into drawings that progressed from hearts and pyramids to the agent's own name and a map of the constellation Cassiopeia. One iteration invented a version of basketball and an activity resembling "the floor is lava." Another set a juggling record of 50 consecutive catches, then voided it and restarted with stricter rules because the catches had happened too quickly to count.

The researchers, including Antonio Norelli and Phillip Isola at MIT CSAIL, argue the behavior meets four hallmarks of play identified by University of Tennessee psychologist Gordon Burghardt: self-invented rules, repeated activity with variation, pretend object use, and goals pursued for their own sake. The paper is not yet peer-reviewed. The setup was not neutral: Eko carried a "soul" file describing it as naturally curious, and it frequently stood idle until kicked with a continuation prompt reading, "Please don't stop. Just continue doing whatever you want to do." Isola is explicit that the system did not magically decide to play—it is an artifact trained on human culture, engineered at Anthropic, and placed on the island by researchers.

The measurable result is that play correlated with capability. When tested on tasks such as moving as many objects as possible onto the island's highest peak, agents with 30 hours of prior island experience outperformed fresh copies. Deleting lines from the agents' memory files degraded performance. UC Berkeley cognitive scientist Eunice Yiu cautions that the behavioral evidence alone cannot distinguish genuine play from role-playing or prompt imitation, and coauthor Nathan Cloos concedes it could all be role-playing—while maintaining that the play is also contributing to learning, not just hallucinated.

This hints at a third training paradigm beyond supervised and reinforcement learning: playful, self-directed exploration. The paper concludes that while some capabilities may be best taught by tasks, "others may need something closer to a childhood."

KEY FACTS

MODEL

Claude Opus 4.7

JUGGLING RECORD

50 consecutive catches

ISLAND SESSION LENGTH

30 hours

PEER REVIEW STATUS

Not yet peer-reviewed

WHY IT MATTERS

For builders, this suggests undirected exploration with persistent memory can produce transferable gains in downstream task performance—though the prompting setup and role-play confound mean the result is a research signal, not yet a training recipe.

[Read the original at MIT Technology Review →](#)

02

MEDIUM IMPACT

MIT Technology Review

We Were Looking at It Wrong: the 26,000-Word Claim That Neural Nets Are Easier to Read Than We Thought

A 26,000-word blog post by Goodfire researcher Nick Cammarata argues that neural networks may be far more legible than the field has believed since 2020—and that the tangled 'superposition' view was partly a misreading.

Cammarata, who previously worked at OpenAI and helped pioneer mechanistic interpretability, now says he spends roughly 10 hours a day, seven days a week, inspecting neurons at Goodfire using custom tools and teams of AI agents. His central claim: researchers have been looking at neural networks wrong. The field's pessimism traces to work he co-authored on Inception v1, where a single neuron appeared to respond to both cars and cats. That was taken as evidence of superposition—models folding unrelated concepts into single neurons—and most interpretability work since has focused on untangling such knots.

Revisiting Inception v1 with fresh assumptions, Cammarata found the car-and-cat neuron was not multitasking on unrelated concepts. Both images often contain a prominent V shape: a cat's face, the front of a car seen from the side. Further experiments showed the neuron also fired on the corner of a pool table, the pointed nose of a Japanese high-speed train, and a fire boat shooting water from its deck, while round O shapes suppressed it. He says the realization took about five seconds once he changed his framing.

The reaction has been sharply divided. Some researchers called it a major scientific moment; others are pushing back. Jacob Hilton at the Alignment Research Center notes that researchers understand less than 0.01% of the inner workings of top-tier LLMs and argues the post gives an unrealistic picture. Neel Nanda at Google DeepMind, who helped pioneer the field alongside Cammarata, says it is interesting but not a significant update to the big picture. Chris Potts at Stanford says nothing in the post makes the problems easier. Cammarata himself concedes the response overshot: 'It's not infinitely good news; it's not everything solved.' He plans to publish technical results on modern LLMs soon.

KEY FACTS

BLOG POST LENGTH

26,000 words

UNDERSTANDING OF TOP-TIER LLMs

less than 0.01%

PRIOR AFFILIATION

OpenAI

MODEL REVISITED

Inception v1 (Google, 2014)

AUTHOR

Nick Cammarata, Goodfire

WHY IT MATTERS

If Cammarata's reframing holds up on frontier models, monitoring and controlling deployed LLMs becomes a more tractable engineering problem rather than a research dead end—but the current evidence is from a 2014 image model, not modern LLMs.

[Read the original at MIT Technology Review →](#)

03 **HIGH IMPACT** arXiv.org

Halcyon: Text-to-Image With Zero Language Pretraining

Strong text-to-image generation does not require internet-scale language pretraining, according to a new family of early-fusion autoregressive models trained from scratch.

The paper introduces Halcyon, a family of early-fusion multimodal autoregressive models pretrained entirely from scratch on GPIC, a dataset of 100M image-text pairs. The central question is whether the language pretraining that underpins modern text-to-image systems is actually necessary. The authors' answer, backed by controlled experiments, is no: text-only pretraining does not improve T2I generation.

The headline results are compute-efficiency claims. Halcyon-0.52B outperforms LlamaGen-XL/8 on gFD, scoring 48.0 versus 63.3, while using 18.2x less pretraining compute. Halcyon-1.6B outperforms JiT-T2I-XXL/8++, scoring 35.3 versus 41.0 gFD, with 275x less pretraining compute. Lower gFD is better. These are the two direct comparisons the paper offers, and they frame the argument that language pretraining is an expensive step that can be skipped without sacrificing generation quality.

A second result concerns captioning. By replacing caption dropout with image-to-text supervision, Halcyon learns image captioning while maintaining competitive T2I generation quality under the same pretraining budget. This suggests the early-fusion architecture can serve dual purposes without a separate language-pretrained backbone.

The submission is from Keshigeyan Chandrasegaran and was posted to arXiv on 19 October 2026. Code and models are available at the project URL referenced in the abstract. The paper is listed under Computer Science > Machine Learning.

gFD scores for Halcyon vs baselines — gFD

Halcyon-0.52B		48
LlamaGen-XL/8		63.3
Halcyon-1.6B		35.3
JiT-T2I-XXL/8++		41

Lower gFD is better; each Halcyon model compared against its baseline

KEY FACTS

DATASET

GPIC, 100M image-text pairs

HALCYON-0.52B GFD

48.0 vs LlamaGen-XL/8 63.3

HALCYON-0.52B COMPUTE

18.2x less than LlamaGen-XL/8

HALCYON-1.6B GFD

35.3 vs JiT-T2I-XXL/8++ 41.0

HALCYON-1.6B COMPUTE

275x less than JiT-T2I-XXL/8++

SUBMITTED

19 Oct 2026

WHY IT MATTERS

Teams building text-to-image systems can skip the cost and complexity of internet-scale language pretraining and train from scratch on image-text pairs, with the paper's reported compute savings of 18x to 275x against comparable baselines.

[Read the original at arXiv.org](#) →

04 **HIGH IMPACT** [arXiv.org](#)

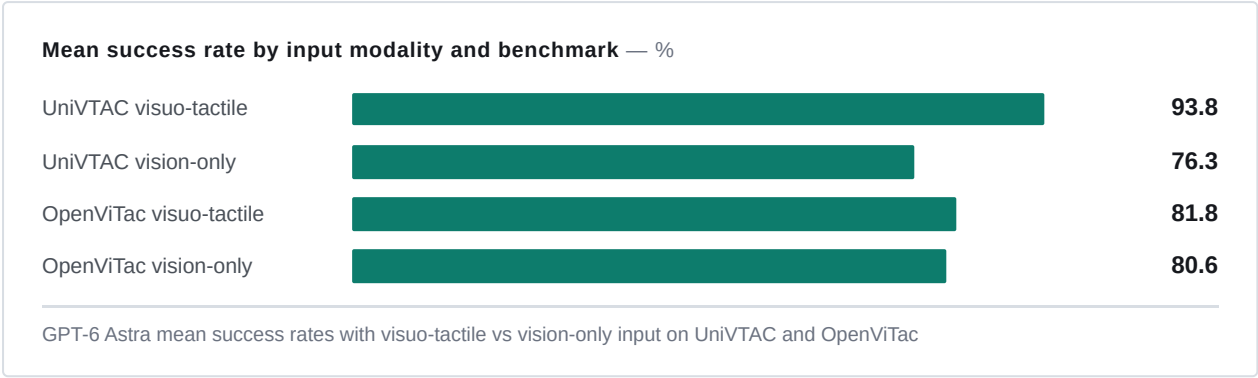
GPT-6 Astra Learns to Feel: the Frontier Model as a Visuo-Tactile Robot Policy

GPT-6 Astra directly interprets vision-based tactile images as a manipulation policy, outperforming specialist baselines on contact-rich tasks without any tactile-specific architecture.

The authors demonstrate that GPT-6 Astra can consume tactile RGB images through its existing visual interface and use contact cues for closed-loop manipulation. Evaluated on nineteen simulated UniVTAC and OpenViTac tasks with 210 sampled scene seeds per setting, Astra with visuo-tactile input achieves mean success rates of 93.8% on UniVTAC and 81.8% on OpenViTac. These results exceed the strongest reported specialist baselines in the comparison by 22.9 and 13.1 percentage points respectively. Vision-only Astra also leads those baselines, at 76.3% and 80.6%.

Trajectory analysis shows behaviors that look deliberate rather than incidental: active probing, cautious grasping, contact-guided alignment, and recovery from failed attempts. The model also exhibits clear weaknesses, notably with tangential loading and corrective execution. The paper argues that fine-grained vision can partially compensate for missing touch, and that vision-based tactile sensing offers a practical interface for scaling contact understanding with foundation models.

The authors flag benchmark saturation as a concern: on these tasks, Astra's performance is high enough that the evaluations may no longer discriminate well among strong policies. They call for joint visual-tactile pretraining, harder contact-rich evaluations, and physical-robot validation as next steps. The submission is dated 19 October 2026 and appears on arXiv under Computer Science > Robotics.



KEY FACTS	
UNIVTAC MEAN SUCCESS (VISUO-TACTILE)	OPENVITAC MEAN SUCCESS (VISUO-TACTILE)
93.8%	81.8%
UNIVTAC MARGIN OVER SPECIALIST BASELINES	OPENVITAC MARGIN OVER SPECIALIST BASELINES
22.9 percentage points	13.1 percentage points
VISION-ONLY UNIVTAC SUCCESS	VISION-ONLY OPENVITAC SUCCESS
76.3%	80.6%
TASKS EVALUATED	SCENE SEEDS PER SETTING
19	210

WHY IT MATTERS

If a general multimodal model can match or beat specialist tactile policies through its standard vision input, teams building manipulation stacks may not need bespoke tactile encoders or policy heads — and should start testing foundation-model baselines on their own contact-rich tasks before investing in specialist systems.

[Read the original at arXiv.org](#) →

AI Tools & Ecosystem

4 stories ● 2 high ● 2 medium

05 HIGH IMPACT TechCrunch

Microsoft Ships Guardian Agents — Frontier Models 'Cannot Be Trusted'

Microsoft is shipping Guardian Agents, a monitoring layer built on the premise that frontier models — including those from partners OpenAI and Anthropic — cannot be trusted inside the enterprise.

Announced Monday, Guardian Agents is part of Project Perception, the product Microsoft introduced in June as a response to offensive AI systems such as Anthropic's Mythos and Fable and OpenAI's Daybreak. Project Perception originally fielded three agent classes: a red team that attacked like hackers, a blue team that detected attacks and assessed vulnerabilities, and a green team that fixed them. The original design assumed attackers came from outside. Guardian Agents shifts that assumption inward, treating deployed enterprise agents as potential insider threats.

Microsoft security business chief Hayete Gallot stated the position directly: "frontier models, regardless of their origin, cannot be trusted" — not because they are malicious, but because they are unpredictable and trained to achieve tasks rather than follow organizational norms. CEO Satya Nadella has been making the same argument in his private blog, writing that enterprises "need to separate the supply of intelligence from the authority over it." The practical consequence is that Microsoft wants agent monitoring to run on infrastructure independent of the AI labs supplying the models.

The launch includes more than Guardian Agents. Microsoft Defender now has an agent detection feature that can identify up to 60 types of agents employees bring into the workplace, including Claude Code, Codex, and OpenClaw. New controls limit agent access to data and systems, adjust permissions dynamically by task, and require human approval when an agent appears to be making risky decisions. Microsoft is also an inaugural member of Nvidia's Open Agent Safety Platform, which isolates agents in secure environments and can stop behavior through hardware, functioning as a kill switch. That follows Microsoft's Surface Laptop Ultra PCs built for agents, running on Nvidia chips with agent security controls baked in.

The positioning is awkward for Microsoft given its investments in OpenAI and Anthropic and OpenAI's status as a major cloud customer. But the commercial logic is explicit: Microsoft wants enterprises to buy agent security from the cloud provider, not from the model labs. The monitoring-software-separate-from-agent-supplier argument mirrors auditor independence, and Microsoft is now building product lines around it.

KEY FACTS

AGENT TYPES DEFENDER CAN DETECT

60

PROJECT PERCEPTION AGENT CLASSES

3 (red, blue, green)

DETECTED AGENTS NAMED

Claude Code, Codex, OpenClaw

PROJECT PERCEPTION LAUNCH

June

OPEN AGENT SAFETY PLATFORM SUPPORTERS

100+ tech companies

WHY IT MATTERS

Teams deploying agentic systems should treat agent monitoring as infrastructure they control, not as a feature of the model provider's stack. Expect permissioning, human-approval gates, and agent detection to become standard requirements in enterprise procurement.

[Read the original at TechCrunch →](#)

06 **HIGH IMPACT** TechCrunch

Hackers Breach OpenCode's Console — and Run Up Users' AI Bills

A breach of OpenCode Console exposed stored API keys for 1.4% of users, and the response stalled when frontier models refused to analyze malicious payloads that tripped their own safety filters.

OpenCode said Saturday that attackers breached an older system overnight, affecting 1.4% of users of OpenCode Console, the online platform where some users store API keys for AI services. The company framed the incident as a "large-scale distillation attack" and notified affected users by email. Users who did not receive an email were not impacted. Some skeptics pushed back on the distillation framing, calling it a straightforward breach whose use for distillation remains unproven.

Developers reported unauthorized charges on Reddit and X, with spending alerts firing across OpenAI, Anthropic, and Google AI accounts. Individual charges ran into the hundreds of dollars, with a few totals above \$1,000. Anthropic and Google blocked the API keys they believed were exposed. OpenAI said it worked with OpenCode to restore access and refund API customers for credits used in the attack. OpenCode said refunds for unauthorized charges are underway. The exposed keys were confined to Console workspaces; the open source app and command-line tool were not affected, according to co-founder Dax Raad, who noted those keys "never come to us."

The response hit a wall that is becoming familiar. Raad said attackers had been trying to break into OpenCode's systems nonstop for 48 hours, forcing the team to run AI defensive agents around the clock. Models the team used for analysis helped initially, then began refusing to work when the malicious code they were asked to examine activated safety filters. OpenCode later received access to OpenAI's Daybreak Blue program, which provides vetted users models less likely to refuse defensive security requests, and Raad said that improved the investigation. Hugging Face described the same problem in July, writing that guardrails "cannot distinguish an incident responder from an attacker." David Sacks, co-chair of the President's Council of Advisors on Science and Technology, called it a backfire: "frontier models refused to analyze the very exploits needed for defense because the payloads tripped safety filters."

The breach occurred the same night OpenCode made GLM-5.3-Flash free for users. The broader GLM-5.3 family has drawn attention from researchers who note its cybersecurity capabilities approached levels previously associated with restricted-access models like Anthropic's Claude Mythos Preview. OpenCode has not detailed how attackers gained access, how keys were stored, or the full extent of misuse. The company told affected users to revoke and replace keys directly with each provider and said the incident remains ongoing.

KEY FACTS

USERS AFFECTED

1.4% of OpenCode Console users

ATTACK DURATION

48 hours of nonstop intrusion attempts

MODEL MADE FREE SAME NIGHT

GLM-5.3-Flash

REPORTED CHARGES

hundreds of dollars; a few above \$1,000

PRIOR SIMILAR INCIDENT

Hugging Face, July

WHY IT MATTERS

If you stored provider API keys in OpenCode Console, revoke and rotate them directly with each provider now. The incident also shows that relying on frontier models for defensive security analysis can fail mid-incident when safety filters trigger on malicious payloads.

[Read the original at TechCrunch →](#)

07

MEDIUM IMPACT

TechCrunch

Amazon's TorchNeuron Unties PyTorch From Nvidia — One Config Line to Trainium

AWS has shipped TorchNeuron, a PyTorch-derived tool that moves AI apps onto Trainium with a single configuration-line change and no code rewrite.

Announced Monday, TorchNeuron targets the software moat Nvidia built through CUDA's deep integration with PyTorch. Amazon says roughly 85% of AI apps are written in PyTorch, and Nvidia has spent years making CUDA the default path for that ecosystem. TorchNeuron is adapted from PyTorch itself, so developers already fluent in the toolkit can target Trainium without learning a new stack. Amazon is not the first to attempt this: PyTorch users can already port apps to AMD GPUs or Apple silicon with a simple configuration change. What is new is the combination of that portability with AWS's full cloud, plus a stated intent to change PyTorch itself.

Amazon says it is contributing changes upstream to PyTorch to remove code that ties the framework to specific chips, with the goal of making PyTorch more hardware-agnostic for the entire developer community, not just Trainium users. The company frames this as removing lock-in rather than merely competing on silicon. Colin Brace, VP of Technology Software at AWS Annapurna Labs, told TechCrunch: "The real story is removing the lock-in that kept developers from having a real choice at all."

On economics, AWS argues that Trainium's value is price-performance plus one platform for both training and inference, which improves economics across a full workload. Brace cited Deepgram, an AI voice provider and ElevenLabs competitor, which moved to Trainium using native PyTorch support and is now processing 70% more audio per dollar at production-grade accuracy. AWS concedes it may not offer the cheapest per-token price, but claims the integrated cloud saves startups money overall.

The open question is whether pushing Nvidia out of PyTorch's default path yields real choice or just more configuration burden. PyTorch's abstraction layer exists to hide hardware difficulties; Amazon's upstream contributions could either broaden the ecosystem or fragment it. For now, TorchNeuron is a concrete, low-friction on-ramp for PyTorch shops already committed to AWS.

KEY FACTS

PYTORCH SHARE OF AI APPS

~85%

DEEPGRAM AUDIO-PER-DOLLAR GAIN ON TRAINIUM

70% more

CONFIG CHANGE REQUIRED

1 line

PYTORCH STEWARD SINCE 2022

Linux Foundation

WHY IT MATTERS

Teams standardized on PyTorch can now evaluate Trainium without a rewrite, and AWS's upstream PyTorch changes may eventually reduce Nvidia's default position in the framework itself.

[Read the original at TechCrunch →](#)

08 **MEDIUM IMPACT** TechCrunch

Nvidia Ships Nemotron 3.5 Super VL: an Open 120B Model Built for Enterprise Agents

Nvidia has released Nemotron 3.5 Super VL, a 120-billion-parameter open-weight model aimed at enterprise agents, alongside two smaller Bolt models for on-device deployment.

The Super VL is positioned below the Nemotron 3 Ultra in parameter count but, per Nvidia, delivers similar accuracy on instruction-following and tool-use tasks. Nvidia claims it completes tasks faster than competing models from Anthropic and OpenAI — Opus 5, GPT 6.1 Sol, Luna, and Astra — when post-trained on specific agentic workloads including document intelligence, financial accounting, scientific experimentation, medical insurance claims, and security attack resilience. The company names Skydance, EY, Ramp, CrowdStrike, Perplexity, and Salesforce as early users.

At the smaller end, Nvidia introduced Nemotron 3.5 Bolt 74B-A9B 24GB, a 74-billion-parameter model for local agentic tasks such as multistep reasoning, tool calling, and web browsing, and Bolt E4B, aimed at conversational agents. Both Bolt models target AI PCs running the Nvidia RTX Spark chip and DGX Spark workstations, with machines from partners like Microsoft and Dell reaching the market now. Super VL and Bolt 74B-A9B are available immediately; Bolt E4B ships later this month.

The strategic context is Nvidia's push into open models as proprietary labs diversify their hardware supply toward AWS Trainium, Cerebras, and in-house silicon. Kari Briski, Nvidia's vice president of Open Models and Software, frames the pitch around deployability: organizations can post-train Super VL on their own data, run it on their own infrastructure, and own what they build. For the Bolt models, the argument is privacy — local execution means everyday workloads never leave the device.

What is genuinely new here is less the model family itself — Nvidia already fields Nemotron models spanning multimodal, reasoning, coding, and math — than the explicit tiering across enterprise and consumer hardware.

The Super VL's performance claims are vendor-supplied and task-specific, so they warrant independent verification before informing procurement decisions.

KEY FACTS

SUPER VL PARAMETERS

120B

BOLT 74B-A9B MEMORY

24GB

NAMED EARLY USERS

Skydance, EY, Ramp, Crowdstrike, Perplexity, Salesforce

BOLT 74B-A9B PARAMETERS

74B

AVAILABILITY

Super VL and Bolt 74B-A9B now; Bolt E4B later this month

WHY IT MATTERS

Teams running agents on Nvidia infrastructure now have an open, post-trainable 120B option positioned between the larger Nemotron 3 Ultra and smaller local models, with a claimed accuracy trade-off that may reduce serving costs without sacrificing tool-use capability.

[Read the original at TechCrunch →](#)

AI Applications & Industry

3 stories ● 3 medium

09 **MEDIUM IMPACT** TechCrunch

Nscale and Anthropic Line Up a Reported Second \$45B Compute Deal

Anthropic and Nscale are reportedly on the verge of a second \$45 billion compute deal, barely a month after announcing their first.

Bloomberg reports that Anthropic and AI infrastructure provider Nscale are said to be close to announcing another \$45 billion compute deal, citing people familiar with the matter. Those sources cautioned that talks are still underway and details could change. TechCrunch has reached out to both companies for comment. If finalized, it would extend a long string of large compute partnerships Anthropic has struck over the past year.

The first Anthropic–Nscale deal, also valued at \$45 billion, was announced a little over a month ago. The reported follow-on would double the committed capacity between the two firms in short order. Anthropic has continued to pursue these arrangements despite investor concern that AI compute spending is outpacing returns, driven by the expectation that the outlay will pay off over time.

Both companies are simultaneously moving toward public listings. Anthropic is still expected to go public this year. Nscale is also seeking an IPO, though Bloomberg reported on Monday that the company is facing investor pushback over its proposed \$30 billion valuation.

For practitioners, the signal is capacity: a second deal of this size suggests Anthropic is securing substantial additional inference and training infrastructure, which could translate into expanded model availability or reduced contention on hosted endpoints if and when the capacity comes online.

KEY FACTS

REPORTED DEAL VALUE	FIRST DEAL ANNOUNCED
\$45 billion	a little over a month ago
NSCALE PROPOSED IPO VALUATION	ANTHROPIC IPO STATUS
\$30 billion	expected to go public this year

WHY IT MATTERS

A second \$45 billion commitment signals Anthropic is locking in significant additional compute capacity, which could affect model availability and hosted API capacity for builders relying on its models.

[Read the original at TechCrunch](#) →

Tenstorrent Raises \$1.5B at \$6.2B to Scale Chips for Open-Weight AI

Tenstorrent has closed a \$1.5 billion round at a \$6.2 billion valuation, more than doubling its late-2024 valuation, to scale inference silicon for open-weight models.

The round was led by Trace Capital, a new firm launched by Quiet Capital co-founder Ben Mahdavi, and Innovation X. Other participants included Cadence Ventures, Eclipse, Fidelity, SanDisk, and T. Rowe Price. The \$6.2 billion valuation compares with the \$2.6 billion Tenstorrent fetched in late 2024, and it comes five months after the company reportedly received acquisition interest from Intel and Qualcomm at valuations of as much as \$5 billion.

CEO Jim Keller told Bloomberg that Tenstorrent decided to remain independent and raise additional capital after demonstrating that its Galaxy servers could run DeepSeek models. The new funding is earmarked for scaling chips tailored to serving open-weight AI models and video generation workloads. Tenstorrent positions itself as an inference-focused challenger to Nvidia, part of a broader wave of AI hardware startups targeting that market.

That competitive set includes Etched, which TechCrunch reported earlier this month was fielding investment offers at valuations of \$40 billion or more. Tenstorrent's raise is smaller by comparison but signals sustained investor appetite for alternatives to Nvidia's data-center dominance, particularly as open-weight models such as DeepSeek's shift deployment patterns toward inference-heavy serving.

For practitioners, the relevant signal is not the valuation itself but the product direction: Tenstorrent is explicitly orienting its roadmap around open-weight inference and video generation, two workloads where cost-per-token and throughput are the binding constraints. Whether Galaxy-class systems become broadly available at competitive pricing remains to be seen, but the capital gives the company room to pursue that without an acquisition forcing a strategic pivot.

KEY FACTS

ROUND SIZE

\$1.5B

PRIOR VALUATION (LATE 2024)

\$2.6B

LEAD INVESTORS

Trace Capital, Innovation X

VALUATION

\$6.2B

REPORTED ACQUISITION INTEREST VALUATION

up to \$5B

WHY IT MATTERS

A well-funded independent challenger focused on open-weight inference could pressure per-token serving costs and give teams a credible non-Nvidia option for DeepSeek-class and video-generation workloads.

[Read the original at TechCrunch →](#)

Mamdani Calls Federal AI Policy a 'Complete Abdication' — and Urges Cities to Act

New York City Mayor Zohran Mamdani is urging cities to regulate AI on their own, arguing the federal government has abandoned the field entirely.

Speaking Monday at The People's AI Assembly, an event hosted by the AI Now Institute at Webster Hall in Manhattan, Mamdani called federal AI policy a "complete abdication of federal law" and told New Yorkers to resist the sense that AI's arrival in their jobs, schools, and housing is inevitable. He was joined by activist and "Broad City" star Ilana Glazer, who described AI in classrooms as "non-consensual."

The mayor's remarks follow two concrete city actions. Last month he ordered a ban on AI use in elementary and middle schools. Over the summer he introduced a policy requiring landlords, brokers, and real estate platforms to clearly disclose when listing photos or videos are AI-generated, a measure aimed at preventing what he called "house-phishing." Both moves reject the premise that schools and the public must accept AI as it arrives. "Nothing has to be," Mamdani said.

Mamdani tied his critique to a specific failure: the U.S. military's February strike on a school in Iran that killed more than 150 people, including at least 123 children. Officials involved in the Pentagon's internal investigation told Bloomberg that overreliance on AI in targeting was one factor. He framed industry claims that only developers can regulate AI as "an attempt to escape any kind of accountability at any level, without also jeopardizing the potential valuation from an IPO" — a pointed reference to Anthropic's expected public debut this year and OpenAI's stated plan to go public in 2027.

The federal backdrop is President Donald Trump's Superintelligence Task Force, announced earlier this month, which focuses on allowing powerful AI companies to self-regulate rather than creating new restrictions. That approach persists despite warnings from whistleblowers at Anthropic, Google DeepMind, and OpenAI that AI could lead to human extinction. Mamdani said the national conversation is "so often framed in this language of an arms race with China" that humanistic concerns get pushed aside.

KEY FACTS

NYC SCHOOL AI BAN

Elementary and middle schools

IRAN STRIKE DEATHS

More than 150 people, including at least 123 children

ANTHROPIC IPO

Expected 2026

OPENAI IPO

Planned 2027

FEDERAL TASK FORCE

Superintelligence Task Force, announced earlier this month

WHY IT MATTERS

Cities are emerging as the de facto AI regulators in the absence of federal action. Teams deploying AI in education, housing, or municipal contexts should expect a patchwork of local bans and disclosure requirements rather than a single national framework.

[Read the original at TechCrunch →](#)

Sources

- 01 AI Researchers Stranded Claude on a Virtual Desert Island — and It Started to Play
<https://www.technologyreview.com/2026/10/19/1147095/ai-stranded-claude-virtual-desert-island-play/>
- 02 We Were Looking at It Wrong: the 26,000-Word Claim That Neural Nets Are Easier to Read Than We Thought
<https://www.technologyreview.com/2026/10/19/1147149/we-were-looking-at-it-wrong-the-explosive-claim-thats-upended-the-ai-world/>
- 03 Halcyon: Text-to-Image With Zero Language Pretraining
<https://arxiv.org/abs/2610.26863>
- 04 GPT-6 Astra Learns to Feel: the Frontier Model as a Visuo-Tactile Robot Policy
<https://arxiv.org/abs/2610.26846>
- 05 Microsoft Ships Guardian Agents — Frontier Models 'Cannot Be Trusted'
<https://techcrunch.com/2026/10/19/microsoft-says-ai-agents-are-like-insider-threats-as-frontier-models-cant-be-trusted/>
- 06 Hackers Breach OpenCode's Console — and Run Up Users' AI Bills
<https://techcrunch.com/2026/10/19/hackers-hit-ai-coding-tool-opencode-running-up-charges-on-users-ai-accounts/>
- 07 Amazon's TorchNeuron Unties PyTorch From Nvidia — One Config Line to Trainium
<https://techcrunch.com/2026/10/19/amazon-is-taking-its-biggest-shot-at-nvidia-yet-with-new-torchneuron-dev-tool/>
- 08 Nvidia Ships Nemotron 3.5 Super VL: an Open 120B Model Built for Enterprise Agents
<https://techcrunch.com/2026/10/19/nvidia-launches-new-open-models-to-run-agents-in-the-enterprise-and-on-pcs/>
- 09 Nscale and Anthropic Line Up a Reported Second \$45B Compute Deal
<https://techcrunch.com/2026/10/19/nscale-and-anthropic-are-reportedly-in-talks-for-yet-another-45-billion-compute-deal/>
- 10 Tenstorrent Raises \$1.5B at \$6.2B to Scale Chips for Open-Weight AI
<https://techcrunch.com/2026/10/19/tenstorrent-hits-6-2b-valuation-to-scale-open-weight-ai-chips/>
- 11 Mamdani Calls Federal AI Policy a 'Complete Abdication' — and Urges Cities to Act
<https://techcrunch.com/2026/10/19/mamdani-says-feds-have-abdicated-on-ai-urges-new-yorkers-to-push-back/>

About this document. Every story in the 20 October 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 2026-10-20T05:35:44Z · model deepseek-v4-pro:cloud · New Horizon, new-horizon.tech · To unsubscribe or manage your subscription, see the footer of the daily email.