

Meta releases Muse Spark Open weights

Spyware through targeted ads, METR inside Anthropic and \$15 million to help mathematicians adapt to AI

Friday, 16 October 2026

IN THIS ISSUE

1. Meta opened the weights of Muse Spark Open, the model runs on two Mac Studios
2. Targeted ads delivered spyware without a single click
3. Anthropic let METR inside with employee-level access
4. Mathematicians got \$15 million to adapt to AI and write about grief
5. Goodfire: neural networks are much easier to understand than the field thought
6. Stocks recovered, but insurance against AI company defaults becomes more expensive
7. Accenture limited Claude for employees to \$29 per month
8. Sites already write instructions for agents in HTTP headers
9. Mononoke author: the copy of Scion was deliberate, Anthropic added credit
10. Claude helped write a driver for undocumented hardware in one day

TOPIC 1

Meta opened the weights of Muse Spark Open, the model runs on two Mac Studios

SOURCES Meta AI, blog, October 15 [Meta](#) · Hugging Face, blog, October 15 [Hugging Face](#) · Artificial Analysis, October 15 [Artificialanalysis](#) · The Wall Street Journal, October 15 [WSJ](#) · Aaron Levie, X, October 15 [@levie](#) · HN discussion, October 15 [Hacker News](#)

On Tuesday this digest covered Muse Wonder, the closed model of Meta. On October 15 the company released the weights of another model in this series, Muse Spark Open. Meta calls it the strongest American model with open weights. It is an MoE with 816 billion parameters, and 42 billion of them are active for each token. The context is up to 1 million tokens. The model accepts text, images, audio and video, and it answers with text.

Hugging Face described the architecture in detail. Each of the 62 layers has 256 experts, and the router selects 8 for each token. The checkpoint in BF16 is about 1.5 TB. The full context of 1 million tokens takes about 266 GB in the cache. Transformers, vLLM, SGLang, llama.cpp and MLX already support the model.

The main trick of the release is a launch at home. The WSJ writes about a partnership of Meta and Apple. The M5 Ultra in a Mac Studio gives the GPU up to 512 GB of unified memory. Thus two machines together hold the full model. The head of Hugging Face,

Clement Delangue, already showed a quantized version on two Mac Studios. The Muse Model License permits commercial use with restrictions.

The independent evaluation is more modest than the Meta claim. Artificial Analysis gives the model 39 points in its index. This is the level of DeepSeek V4.1 Flash (39), but below MiMo-V2.6-Pro (46)

and GLM-5.3 (45). The API price is \$1.25 per 1 million input tokens and \$4.25 per 1 million output tokens. For an open model of this size, this is expensive. On HN, several developers write that smaller Chinese models give the same result at a lower price.

Why it matters. The title "best American open model" has weight for policy and corporate procurement, because some companies do not take Chinese weights. Technically, China still leads in open models. The new part is the hardware. A model at the frontier level of a year ago now runs on two desktop computers without a cloud.

TOPIC 2

Targeted ads delivered spyware without a single click

SOURCES Citizen Lab, report, October 15 [Citizenlab](#) · Ars Technica, October 15 [Ars Technica](#) · The Guardian, October 15 [Guardian](#) · HN discussion, October 15 [Hacker News](#)

Citizen Lab at the University of Toronto described the AdFection campaign. According to the lab, this is the first confirmed infection with commercial spyware through targeted ads. The victim does not need to click anything. It is sufficient to open a site or an app that shows ads. The attack worked on iOS and Android. Citizen Lab confirmed 45 infected devices at Al Jazeera in 2024-2025.

The Israeli company Discover-Tech, which calls itself an ad agency, built the scheme. Its platform Controlium collects profiles of people from ad auctions: device ID, IP, geolocation. When a target matches a profile, the ad loads code from Candiru servers. Candiru now operates under the name Integrity Labs. Its server sends the exploit, and the device becomes infected in seconds.

The scale is large. Citizen Lab obtained more than 4 billion Discover-Tech bid records. From March to June 2026, its ads went to more than 70,000 apps and 120,000 sites. Impressions occurred in the networks of NATO, the UN, European parliaments and the armies of Norway and Sweden. Google and IAB Europe had Discover-Tech on their lists of verified vendors, and after the report Google removed the company from its ad exchanges. Discover-Tech and Candiru did not answer requests.

Why it matters. An ad auction sends the profile of a person to hundreds of companies in milliseconds. Until now, people saw this as a privacy problem. The report shows that the same channel delivers an exploit to a specific person. For newsrooms, governments and companies, an ad blocker becomes a security tool. Citizen Lab also advises to reset the advertising ID in the phone settings.

TOPIC 3

Anthropic let METR inside with employee-level access

SOURCES Anthropic, October 15 [Anthropic](#) · Anthropic, X, October 15 [@AnthropicAI](#) · Financial Times, October 15 [FT](#) · Ars Technica, October 15 [Ars Technica](#)

On September 13 this digest covered the essay by Dario Amodei. At that time Anthropic only promised to let outside evaluators inside. On October 15 the company reported that six METR staff members already have work hardware, an office and access to internal systems. The level of access is approximately the same as for its own risk assessment teams. Only a few categories stay closed:

customer data, attorney-client privilege, model architecture.

METR will publish conclusions about the level of risk, incidents and company practices. Anthropic cannot block a publication. It can only demand the removal of information from narrow categories.

METR has the right to say publicly if it considers these edits justified. The company compares the scheme to bank supervision: the regulator puts its inspectors inside a large bank.

For other partners, Anthropic builds three levels of access. SecureBio now checks the section on chemical and biological weapons in the August risk report. With Accenture and Faculty, the company plans a wider cooperation, with red teaming and alignment evaluation. Hugging Face is among the organizations where talks are still at an early stage.

In the same week, according to Axios as retold by Ars, the FTC sent Anthropic and OpenAI demands for documents. The request contains more than 50 pages of questions about recent incidents with agents and about safety measures. Ars reminds that METR will probably also get a subpoena.

Why it matters. This is the first case when a lab gives an outside organization such access and the right to publish without a veto. The test will be the first METR report: what it says and how much it hides. There is also a weak point. METR has a long history of work with Anthropic. Thus the independence here rests only on a contract.

TOPIC 4

Mathematicians got \$15 million to adapt to AI and write about grief

SOURCES Simons Foundation, October 15 [Simonsfoundation](#) · OpenAI, X, October 15 [@OpenAI](#) · Anthropic, X, October 15 [@AnthropicAI](#) · Derek Thompson, October 15 [Derekthompson](#) · Rikab Gambhir on the blog of Terence Tao, October 15 [Wordpress](#) · HN discussion, October 15 [Hacker News](#)

On October 7 OpenAI released 722 mathematical papers that its models wrote. A week later the reaction of the community became money and essays. The Simons Foundation, together with Anthropic and OpenAI, gives \$15 million. OpenAI and Anthropic give \$5 million each.

The foundation will fund workshops and projects: verification of results with AI, training on new tools, common standards.

The president of the foundation, David Spergel, says that the community of mathematicians must set the rules itself.

The essay by Derek Thompson "I know why the mathematicians weep" got 575 comments on HN. He quotes the Fields medalist Hugo Duminil-Copin: "They destroyed my field of research." Steven Strogatz of Cornell says that colleagues speak about "open-pit mining" of their subject. Thompson explains the grief like this: science is a craft and a conversation. AI does the craft well, but it breaks a conversation between people that is thousands of years old.

On the blog of Terence Tao, the physicist Rikab Gambhir published a guest post. It is satire. He writes that OpenAI gives mathematicians the most boring work: to check and to understand the results. The results themselves appear faster than anyone can read them.

Why it matters. Mathematicians now have many results. They lack people who can check and understand these results. Money from the same companies that created the flow partly admits this.

Other fields where AI generates results at scale will meet the same question: who checks?

TOPIC 5

Goodfire: neural networks are much easier to understand than the field thought

SOURCES Nick Cammarata, Goodfire, October 15 [Goodfire](#) · The Rundown AI, X, October 15 [@TheRundownAI](#) · HN discussion, October 15 [Hacker News](#)

Nick Cammarata worked on interpretability at OpenAI before, and now he is at Goodfire. He published an essay that takes 65 minutes to read. The main thesis: most neurons in models have one clear role.

For years the field thought otherwise. The explanation had the name superposition: the network supposedly packs unrelated concepts into one neuron because it lacks space.

Cammarata checked again the canonical examples that he himself published in the past. The famous "cats and cars" neuron in InceptionV1 is in fact a detector of a V shape. The circuit where a car "hid" among dogs in fact grouped them by pose. In 2023, an automated OpenAI project found 1.7% of neurons interpretable. A preliminary Goodfire audit of Qwen3-32B gives this picture. About 90% of neurons have one main role, ~7% have two related roles, and ~2-4% are still unclear.

Why it matters. If the thesis holds, interpretability becomes a normal science on the network itself. Then a separate decomposition tool such as a sparse autoencoder is optional. This is directly relevant to safety: a check of a model from the inside becomes

more realistic. For now this is an essay and a "working hypothesis", and the technical paper is still unpublished.

TOPIC 6

Stocks recovered, but insurance against AI company defaults becomes more expensive

SOURCES Semafor, October 15 [Semafor](#) · Semafor, October 15 [Semafor](#) · Financial Times, October 15 [FT](#) · Financial Times, October 15 [FT](#)

Yesterday this digest covered how banks look for buyers of data center debt. On October 15, US stocks recovered part of the losses after several days of sell-off. But the credit market moves in the other direction. Semafor writes that investors buy credit default swaps on hyperscalers in large volumes. Protection for the largest borrowers became the most expensive.

One year ago, insurance on \$1 million of CoreWeave bonds for five years cost less than \$40 thousand per year. Now it costs \$100 thousand. The yield on CoreWeave bonds is above 15%. Protection against an Oracle default became 12% more expensive in a week, to \$28 thousand. The Oracle debt of \$160 billion is at the edge of a "junk" rating. In its headline the FT writes that investors dump data center bonds, and it calls this a "Darwinian moment".

Against this background, TSMC reported a jump in profit. The company says that it builds plants at an "unprecedented pace" because of "extremely strong" demand for AI.

Why it matters. Stocks react to the mood of the day, and the price of insurance shows how the market rates the real risk. The gap between them grows. Chip makers earn record profits, but creditors already fear the companies that build data centers on debt. The weakest links here are companies with large debt and one type of customer.

TOPIC 7

Accenture limited Claude for employees to \$29 per month

SOURCES 404 Media, October 15 [404media](#) · Siqi Chen, X, October 15 [@blader](#)

404 Media saw an internal announcement of Accenture. From October 1, the company pays centrally only for \$29 of Claude use per person per month. A specific project must pay for all use above this amount. Some employees spent \$500-1,000 per month.

One of them writes that he uses Claude Code intensively for an "AI accelerator" for clients. Another employee lost access completely. On the premium plan the limit is different: \$250 of tokens per week. Accenture says that it provides "the right tool for the right job".

The irony is that Accenture advises clients to invest in AI. On the same day, Anthropic named it among its outside safety evaluators (item 3). Siqi Chen reacted: it would be more

correct to give the full budget to the 1% most active users, and nothing to the rest. [single source]

Why it matters. Consumption by AI agents grows faster than budgets. Large companies move from unlimited licenses to payment for actual use. An average limit for all hits exactly the people who get the most from the tool. Every large company will now meet this contradiction.

TOPIC 8

Sites already write instructions for agents in HTTP headers

SOURCES Signals Corps, October 15 [Signalscorps](#) · HN discussion, October 15 [Hacker News](#)

Signals Corps checked the headers of 217 million responses from the September Common Crawl, which is 10% of the full crawl. 7,959 sites sent a header addressed to AI. Almost all of them only point the agent to llms.txt or an MCP server. 30 sites tried to give the agent commands, 24 of them in the X-AI header. Most of these attempts are jokes. Real harm is possible only when the agent has the necessary rights, for example shell or database access.

Another 139 sites wrote into a header whether they consider the client an AI bot. Thus servers already tell an agent from a person and can show them different content. The authors advise to watch the actions of the agent after the request. A filter by header names would give 910 false alarms.

Why it matters. An agent reads more than the text of a page. Headers that a person never sees also go into its context. Any address that a server suggests must be a new untrusted request for the agent.

TOPIC 9

Mononoke author: the copy of Scion was deliberate, Anthropic added credit

SOURCES Rain (sunshowers.io), October 15 [Sunshowers](#) · HN discussion, October 15 [Hacker News](#)

Yesterday this digest covered the post by one of the authors of Mononoke. She compared Anthropic's Scion with the version control server of Meta. In a second post she goes further: after the first post, she privately got confirmation that the copy was deliberate. According to her, the Scion and Mononoke teams share some people. The similarity covers the full data model. Even the name "Scion"

rearranges the syllables of "Bonsai", the Mononoke name for change sets.

She again says that the problem is attribution. It is acceptable to copy good ideas. But the story "our model built and launched the system in 5 months" sounds different if a Meta team built it for 10 years. In an update to the post, she writes that Anthropic answered on HN and added sufficient credit to the blog.

Why it matters. A story about what a model built now has commercial weight. Thus the question "what existed before it" becomes part of the check of such claims. Anthropic corrected the text in one day, but only after public pressure.

TOPIC 10

Claude helped write a driver for undocumented hardware in one day

SOURCES Christian Schaller, GNOME blog, October 15 [Gnome](#) · HN discussion, October 15 [Hacker News](#)

Christian Schaller of Red Hat wanted to scan old negatives. His Plustek scanner got a new chipset, and after that it did not work on Linux. He gave Claude a virtual machine with Windows and the official driver. The Linux host recorded the USB traffic through usbmon. The model read thousands of lines of the dump, searched for patterns and suggested what each byte means. The scanner started to work.

Then Schaller used the same setup for the MIPI camera in his laptop. In one day he got patches for the kernel and libcamera. Now he published a general version, hw-harness, for Claude and Codex.

Why it matters. Before, a driver for an undocumented device required weeks of manual dump analysis. The model does this boring part, and a person checks the result on real hardware. Vendors often do not write drivers for Linux. Thus this method can change device support there.

misc

- **NVIDIA (@NVIDIAAI)** released the open model Nemotron 3.5 Super VL: 120 billion parameters, 12 billion active. It works with text, images, documents and video for agents. [@NVIDIAAI](#)
- **Black Forest Labs (@bfl_ai)** opened the weights of FLUX 3 Image Klein: 5.6 billion parameters, one consumer graphics card is sufficient, native 4K. [@bfl_ai](#)
- **Krea 3:** the image edit model makes an edit in less than 8 seconds for \$0.06. According to the chart from Krea itself, Nano Banana Pro needs 29 seconds and \$0.135. [single source] [@TheRunDownAI](#)
- **OpenAI (@OpenAI)** released GPT-Live-1.1 for voice in ChatGPT and Codex: new voices and better accents in six languages. [@OpenAI](#)
- **OpenAI (@OpenAI)** opened GPT-Rosalind to academic biology labs in the US: up to 12 months of sponsored access. [@OpenAI](#)
- **Steam** updated its client one day after the launch of the new Epic launcher. A cold start now takes 5.23 seconds, and Steam is again faster than Epic.

Respawnfirst

- **Tesla Roadster:** Tesla showed the new Roadster in Texas. The car drove on a vertical ramp, made a short flight on compressed air and accelerated to 96 km/h in 0.81 seconds. Tesla did not name prices or delivery dates. Ethan Mollick (@emollick) compared the demo to an iPhone presentation and reminded that the path from demo to product often takes years.

Theverge · @emollick

- **The second becomes an hour:** at a conference of the International Bureau of Weights and Measures, delegates from 69 countries voted for a leap hour. It will replace the leap second.

Programmers who caught bugs in minutes with 61 seconds can relax. [Guardian](#)

- **Withings:** from October 13, new accounts must pay \$99.99 per year for Withings+ to sync data with Apple Health. A user complaint on the support forum reached HN.

[Withings](#)

- **Current Affairs** prohibited any text and illustrations that AI generates, under a new contract with its union. [Currentaffairs](#)