

Anthropic pitches Congress its own AI safety bill

OpenCode key leak, the Starshield limit in Ukraine and a new open DeepSeek model

Sunday, 18 October 2026

IN THIS ISSUE

1. Anthropic shows members of Congress its own bill on AI safety
2. OpenCode: attackers got the OpenAI and Anthropic keys of users
3. Zelensky: the Starshield limits in Ukraine are a planned operation
4. DeepSeek released V4.1 Flash 1017 under MIT and opened its search
5. Anthropic pays Blackstone more than 10% for debt for Google chips
6. Replit selected OpenAI as its main partner for price and quality
7. Amazon rates the union risk at warehouses with an internal AI tool
8. Chollet and Levie argue about whether programmers will still have work

TOPIC 1

Anthropic shows members of Congress its own bill on AI safety

SOURCES POLITICO, 17.10 [Politico](#) · Zvi Mowshowitz, Substack, 17.10 [Substack](#)

Politico got a confidential draft of the law. According to three sources, Anthropic representatives quietly show it in Congress. The company confirmed the substance. A spokesperson said that the law repeats its Advanced AI Framework, which Anthropic has promoted since June.

The main point of the text is when the government can stop a model. For this, there must be a foreseeable and imminent risk of harm above \$1 billion, or of death or serious injury to more than 50 people. The risk must come from biological weapons, a cyberattack or the loss of developer control over the model. Only the Department of Commerce can request a stop. A federal court in the District of Columbia makes the decision. No other agency can pause the development, the launch or the internal use of a model.

The law gives obligations to labs. They write safety frameworks and publish transparency reports for each new model. They also protect whistleblowers and report critical incidents. Independent evaluators register with the Department of Commerce. A lab can pay them if the payment does not depend on the result. An evaluator reports an "imminent risk" within 72 hours. In exchange, the law takes from the states the right to regulate risk disclosure, third-party evaluations and incident reports.

The reaction from safety advocates is sharp. Mowshowitz called the draft "unacceptable". In his view, it is weaker than the bipartisan FRONTIER Act from the summer. Charlie Bullock saw that the law gives almost no rulemaking power and does not license evaluators. Without licenses, evaluators will compete for the easiest audit. Nathan Calvin writes that the law also takes from state courts the general right to stop a model. Kelsey Piper explains the motive with the experience of the company itself. In the summer the government shut down two Anthropic models with export control after a few phone calls.

Why it matters. This is the first known bill that a frontier lab wrote itself. The company that speaks most loudly about AI risks proposes a higher threshold for state action than the current one. A reader who follows regulation gets two signals here. Federal preemption of state laws becomes the main bargaining chip. And labs fear an arbitrary decision of the executive branch more than weak rules.

TOPIC 2

OpenCode: attackers got the OpenAI and Anthropic keys of users

SOURCES [OpenCode, X, 17.10](#) [@opencode](#) · [HN discussion, 17.10](#) [Hacker News](#)

On the evening of 17.10 OpenCode reported a breach. The company called it a "large-scale distillation attack". The breach affected 1.4% of the users of the OpenCode Console service. The incident continues. The company works with Anthropic, OpenAI and Google to repair the damage.

The email to users is more exact than the post. The attackers got access to an old system that OpenCode was about to retire. This system kept the keys of third-party providers that users added to their workspace. OpenCode asks users to revoke these keys in the provider dashboards and to check recent costs. Under the post, users write that \$100-\$138 disappeared from their accounts in a few minutes.

On HN the top comment explains the confusion about the name. OpenCode Console is a hosted gateway to models, similar to OpenRouter. It has no relation to the OpenCode CLI agent or to the desktop app. A person who runs only the agent locally with personal keys did not send these keys to the server.

Several commenters doubt the word "distillation". The company did not yet show proof that the keys went to collect answers for the training of another model.

Why it matters. A service that keeps the API keys of other people concentrates the risk. One breach gives access to the accounts of thousands of people at the same time. A provider key is safer only where it is necessary, with a cost limit and with a separate key for each service. Then a leak from one service costs one key.

TOPIC 3

Zelensky: the Starshield limits in Ukraine are a planned operation

SOURCES Financial Times, 17.10 [FT](#) · Ukrainska Pravda, 17.10 [Com](#) · Politico Europe [Politico](#) · Kyiv Independent [KyivIndependent](#) · Kyiv Independent on the Telegraph report [KyivIndependent](#)

On 14.10, Starshield terminals, the military version of Starlink, started to lose the connection in motion. The Ukrainian army got no warning. The commander of a strike drone unit told Politico Europe about a limit of 100 km/h on all terminals. His unit stopped all medium-range strikes. "This is comparable to a nuclear strike against us," he said. Interceptor drones that protect cities from Russian drones also suffered.

A day later the failures stopped, but the 160 km/h limit stayed. Soldiers told the Kyiv Independent that this speed is not enough for an interceptor. On 17.10 Zelensky said that Ukraine tested the terminals in the EU. There they worked with no limits. Thus he called the limit a "planned operation".

Nobody takes responsibility. Musk writes that Starshield belongs to the US government, and the army sets the limits, and finally the president. The Telegraph cites four sources and writes that Musk himself proposed the idea and Trump agreed. On 16.10 Trump said that it was not him and, in his opinion, not Musk. The Pentagon is silent. The Republican senator Thom Tillis asks the Inspector General of the Department of Defense for an investigation.

Why it matters. The connection for drones is now part of the defense capability. This case shows that a supplier can change it with one setting and with no warning. Zelensky says that Europe made its conclusion and already works on an alternative to Starlink.

TOPIC 4

DeepSeek released V4.1 Flash 1017 under MIT and opened its search

SOURCES DeepSeek, API news, 17.10 [Deepseek](#) · Hugging Face, model card [Hugging Face](#) · HN discussion, 17.10 [Hacker News](#)

DeepSeek calls the release a small update of V4.1 Flash, which came out on 10.09. The weights are on Hugging Face under the MIT license. The model page gives 763 billion parameters and a context of 1 million tokens. The company promises better results, faster task execution and shorter answers.

The table in the model card is interesting because of the comparison with closed models. On DeepSWE v1.1 the new model has 77.0. GPT-5.6 Sol has 73.0, and Opus 5 has 74.0. On Automation-Bench it has 60.5 against 50.3 for Opus 5. On Terminal-Bench 4.0 Opus 5 stays ahead: 51.8 against 42.3. DeepSeek itself gave all the numbers. It ran Terminal-Bench in the Claude Code harness.

The second news item in the same post is search. The Search API and the Fetch API are now open for all, at \$0.75 for 1000 calls. The search works on the own search engine of DeepSeek. Calls through the server tools in Claude Code and Codex are free until 31.10.

On HN people praise this model for its predictability. One commenter writes that it generates the code that a developer would write, and it does not try to be too clever.

Another one writes that he regularly runs reviews on five open models, and DeepSeek does the best job.

Why it matters. An open model again reached the leaders of last year on agent tasks. For a team that counts its token costs, this is a strong argument. But for now the numbers come only from the vendor. Thus a test on a team's own task set is necessary before a model change.

TOPIC 5

Anthropic pays Blackstone more than 10% for debt for Google chips

SOURCES Financial Times, 16.10 [FT](#) · Reuters, 17.10 [Reuters](#) · Mint, 17.10 [Livemint](#)

Yesterday this digest wrote about the BB+ rating and saw only a headline about Blackstone. Now there are details. According to the FT, Anthropic collects \$60 billion in the Project North Star package.

The money will go to about 1.5 GW of Google tensor processors. Broadcom guarantees about \$42 billion of senior debt.

Blackstone funds must put \$9 billion into junior debt. Anthropic payments for a five-year chip lease secure this debt. This is half of the junior part, which has \$18 billion in total. The coupon is 10%, and with the discount and the fees the yield is even higher. The FT gives a comparison: a borrower with such a rating pays about 7% on the public market. Four months ago similar Anthropic debt cost 8.5%.

Why it matters. The price of money for an AI lab grows faster than the lab itself. The difference between 8.5% and more than 10% in four months shows that investors want a larger risk premium. This risk relates to the volume of infrastructure costs. Reuters writes separately that it could not verify the FT report.

TOPIC 6

Replit selected OpenAI as its main partner for price and quality

SOURCES Amjad Masad (@amasad), 17.10 [@amasad](#) · Samuel Spitz (@samuel_spitz), 17.10 [@samuel_spitz](#) · TechCrunch, 16.10 [Techcrunch](#)

At TechCrunch Disrupt the Replit CEO Amjad Masad said that OpenAI is now the "preferred partner" of the company. After a wave of questions, he gave more detail on X. Replit stays independent of models, and a user can select Claude. But Replit now builds the base of its agent on OpenAI models.

According to him, the reasons are performance, price and partnership. Samuel Spitz from Replit added that on most tasks OpenAI models give the best ratio of price and result, also against open models.

On the same stage Masad described a "self-driving company". An agent first examines each support request. It looks at GitHub issues, the code and the tickets in Linear. A human gets the ready context and decides what to do. An agent also writes the patch in one click. Separately, Replit launched Bring Your Own Cloud: a company can deploy the platform in its own Azure, Google Cloud or AWS.

Why it matters. Large platforms for programmers now select a model by the price per task, and loyalty to one lab becomes weaker. This is good for users: the competition between labs goes into the price. For labs, this is a signal that a benchmark lead without a good price does not keep a customer.

TOPIC 7

Amazon rates the union risk at warehouses with an internal AI tool

SOURCES [openDemocracy](#), 16.10 [Opendemocracy](#) · [HN discussion](#), 17.10 [Hacker News](#)

Documents from the court case of Amazon against a Canadian union describe the Atlas tool. This is an internal AI forecast model. It estimates how probable it is that warehouse workers will form a union, start a sit-down strike or cause another "disruptive labor event". The documents look like training materials for managers in Canada.

Atlas collects "hundreds of data points". Among them are Reddit posts, worker answers in internal surveys and the history of past actions. Each warehouse gets a score from 100 to 900. A lower score means a higher union risk. For the management, a higher score is "better". The warehouses fall into five levels. 16 warehouses in the US, Canada and the UK have the highest risk level, and two of them are British.

Amazon denies the surveillance of individual people. According to a spokesperson, Atlas works only at the warehouse level and helps to see where workers need support.

Why it matters. The same type of model that forecasts demand can easily forecast the behavior of people. The data here is usual: surveys, forums, incident history. The purpose of use sets the boundary between analytics and surveillance. The documents show this purpose directly.

TOPIC 8

Chollet and Levie argue about whether programmers will still have work

SOURCES [François Chollet \(@fchollet\)](#), 17.10 [@fchollet](#) · [Aaron Levie \(@levie\)](#), 17.10 [@levie](#)

François Chollet listed his observations about the "full" automation of programming. There are more programmers now than in 2025, and the demand grows. They work more intensely than ever. And they almost do not write code: agents write about 100% of it. The fourth observation is the sharpest. The profit of the companies where these engineers and agents work usually does not grow from AI.

Chollet gives a system explanation: the automation of one link moves the bottleneck to another link.

Aaron Levie from Box developed the idea. In his view, the main error of the forecasts was the assumption that the world needs a fixed quantity of software. Agents can operate a computer and process data. Thus tasks that people did not digitize before now become software. A bank, a laboratory and a factory now have tasks similar to engineering tasks.

In the replies Levie got the main objection. The demand for software development and the demand for human developers are two different variables. The demand for grain can explode, and the number of farmers will not grow because of it.

Why it matters. The argument moved. Nobody asks now whether agents write code. The question now is who checks and directs the agents, and how many such people are necessary. For an engineer the practical conclusion is simple: the value moves from code to the task statement and the check.

misc

- **The White House** launched the [si.gov](#) site about "Super Intelligence". It collects the Genesis Mission for 15 federal agencies, faster permits for data centers and a program for the export of the American stack. [Si](#)
- **Michael Sellitto (Anthropic)**: the [@MTSlive](#) account writes that the Anthropic head of global affairs told the IMF this: next year Claude can do 80-90% of frontier model training, against 3% in April. [single source] Other sources did not confirm this number yet. [@levie](#)
- **Andrew Bailey**, the Governor of the Bank of England, said at the G30 that the problems of AI agents "must be solved". This continues item 10 from yesterday: large Wall Street banks already restrict agent access to customer accounts. [FT](#)
- **Linux kernel 7.2**: LWN counted 16,418 commits from 2,652 developers, a record number of authors. Among the main contributions is a long list of bug fixes that an LLM wrote. The share of commits with the Reviewed-by tag falls, and now it is 48%. [Lwn](#)
- **Codex Desktop** on Mac writes Crashpad dumps all the time. For the author of the issue, the folder grew to 4.9 GB and 54,504 files in one day, but the app works normally. [GitHub](#)
- **Palantir**: through freedom of information requests, the Guardian found that many NHS trusts in England reject the £330 million Palantir data platform. Some of them call it "duplicative". [Guardian](#)
- **MapQuest** added a layer with surveillance cameras and license plate readers: Flock, Axon, Verkada and others. [Engadget](#)
- **Estonia** keeps the code of its internet voting system on GitHub for public review. According to the authors, it is the same code that runs in the elections. [GitHub](#)

- **Llmspeed** shows how fast a local LLM will run on specific hardware, before the purchase. The numbers come from community benchmarks. [Llmspeed](#)