AI DIGEST

SpaceX catches Starship with its tower after orbit

SpaceX caught a Starship upper stage with its tower, and WSJ says Nvidia financed customers that bought its GPUs.

Sunday, 15 November 2026

IN THIS ISSUE

1. SpaceX caught a Starship with its tower after a return from orbit
2. Nvidia: WSJ on artificial demand, a director's stock sale and 13 months of silence from Huang
3. Project Zero: details of vulnerabilities will come out after 90 days, and after 30 days from 2028
4. Lovable: in thousands of apps, the database is open without a login
5. SatBench: 63% of 173 benchmarks are "solved", but only 24% by independent runs
6. Fable 5.2 against Opus 5.5 in Claude Code: the same result at 2.8 times the cost
7. Google /teamwork: agents formalized the Weil conjectures in 1.8 million lines of Lean
8. Claude Code Projects: a coordinator gives work to parallel threads
9. Italy fined Meta €805 million for access to photos on Android
10. US ground stop: L3Harris confirmed that it moves FAA communications from copper to fiber

TOPIC 1

SpaceX caught a Starship with its tower after a return from orbit

SOURCES SpaceX, 14.11 Spacex · Ars Technica, 15.11 Ars Technica · @zebulgar, 14.11 @zebulgar · HN, 15.11 Hacker News

On 13.11, Starship lifted off from Starbase in Texas on its 15th flight. The ship went to orbit and released 26 Starlink V3 satellites. It stayed in space for more than 20 hours. Then it lit one Raptor engine and left orbit.

22.5 hours after launch, the ship came back to the same launch tower. Three engines turned it to a vertical position, and the mechanical arms of the tower (Mechazilla) caught it. This is the first such catch of an upper stage.

The Super Heavy booster did not come back this time. Its return burn ended early because of low fuel mass, and SpaceX still looks for the root cause. The flight software sent the booster to an offshore landing in the Gulf.

After the landing burn, the flight termination system fired.

Ars Technica gives the scale. The Space Shuttle made 135 flights in 30 years, and the launch cost was close to $25,000 per pound of cargo. SpaceX now builds one Starship every three

weeks, and the factory is sized for one ship every three days. The main open problem is the heat shield, which must be ready for the next flight fast and cheap.

Why it matters: until now, only first stages of rockets flew again. Now SpaceX has brought both parts of Starship back to the tower, although on different flights. A fully reusable rocket cuts the cost of a launch to orbit many times. This cost decides which projects in space make economic sense at all.

TOPIC 2

Nvidia: WSJ on artificial demand, a director's stock sale and 13 months of silence from Huang

SOURCES TechSpot, 14.11 Techspot · TechPowerUp, 14.11 Techpowerup · The New York Times, 14.11 NYT · Yahoo Finance, 14.11 Yahoo · @garrytan, 14.11 @garrytan · HN, 14.11 Hacker News

Yesterday we wrote about the campaign for the return of Huang. In one day, three new lines appeared.

The first line is business. TechSpot and TechPowerUp retell a WSJ profile of interim CEO Colette Kress. In this profile, Nvidia invested in startups and small cloud providers, which then bought its GPUs. The goal was to sell extra wafers that Nvidia had ordered from TSMC and to keep prices high. In August, Kress proposed to cut such deals, including a $250 billion deal with OpenAI. A WSJ source says that Huang called her a "deer in the headlights". We did not read the WSJ text itself, only the retelling.

The second line is stock. Nvidia director Mark Stevens sold about 4.2 million shares for about $946 million in August and September, Yahoo Finance writes with reference to SEC filings. He had no prearranged trading plan (10b5-1). This alone does not prove a violation, but it adds a question for the board.

The third line is the timeline. NYT published an investigation based on talks with more than 100 people. The paper first asked Nvidia about Alice Huang in September 2025. For 13 months Huang denied a close relationship. He told the company lawyer the truth only last weekend. After the first NYT story, Nvidia stopped the separate report of Singapore sales. Before that, Singapore gave $10 billion, which was 22% of quarterly sales.

The Stand with Jensen campaign continues. Garry Tan of Y Combinator signed the letter and writes about "due process".

Why it matters: the story is now bigger than a personal scandal. If documents confirm the model "we finance a customer so that it buys our chips", the question goes to the whole AI economy. Part of the demand for GPUs could be artificial. The Nvidia quarterly report on 17.11 will be the first point where the company must answer in public.

TOPIC 3

Project Zero: details of vulnerabilities will come out after 90 days, and after 30 days from 2028

SOURCES Google Project Zero, 14.11 Projectzero · HN, 14.11 Hacker News

Since 2021, the Google Project Zero team gave a vendor 90 days for a fix plus 30 days for patch rollout. Now the rule is simple: the details of a vulnerability come out exactly 90 days after the report, with or without a patch. On 02.02.2027 the deadline goes down to 60 days, and on 03.01.2028 to 30 days. The team removes grace periods and other special cases.

Project Zero states the reason directly: LLMs changed the speed of attacks. An attacker with a model finds the same bug faster. From a published patch, the attacker gets the details of the vulnerability almost at once. Thus the team treats the time from patch to exploit as close to zero. One more number: in targets where the team once found a few vulnerabilities a year, it now finds several a week. Talks with each vendor about the status of each bug stopped to scale.

Three exceptions remain: attackers already use the vulnerability, someone else already made it public, or the vendor asks to publish the details.

Why it matters: Project Zero sets an informal standard for the whole industry. When the best known team for vulnerability research says that 90 days is too long, the pressure on vendors grows. For software teams, the result is practical: the time from report to fix becomes a metric that people outside can see. A team needs a process for the case when the details are public and the patch is not ready.

TOPIC 4

Lovable: in thousands of apps, the database is open without a login

SOURCES Crissy Field, 14.11 Crissyfield · HN, 14.11 Hacker News

The German company Crissy Field and the newspaper Handelsblatt checked apps that people built on the Lovable platform. Lovable makes a web app from a description in plain language. The researchers found 159,041 Lovable projects from December 2025 to September 2026 in the Common Crawl archive. 34,545 of them use the Supabase database.

24,664 sites sent database records at once on page load, without a login. With more read-only requests, the researchers got more data. Tables with email addresses were on 7,403 sites, or 30%. The authors say that this is not a confirmed leak rate: some addresses are public or test data.

The serious cases follow. A German lifestyle forum exposed more than 80,000 profiles and more than 200,000 chat messages. A service company exposed more than 1,000 records with passport data. The researchers also found ticket and payment data of event organizers, contracts, invoices and 280,000 emails from one provider to its customers. They could not find out if anyone already used this data.

Why it matters: an access control error can occur in any code. But on a platform where apps follow one template, one error repeats in thousands of places. An attacker needs only one search method. Lovable puts the responsibility for data on the customer in its terms of use. Thus the rule is simple: before an app gets real customer data, a person with security skills must check its access rights.

TOPIC 5

SatBench: 63% of 173 benchmarks are "solved", but only 24% by independent runs

SOURCES @emollick, 14.11 @emollick · SatBench on GitHub GitHub

Ethan Mollick of Wharton built SatBench on the data of the Benchmark Almanac. He took the 173 benchmarks that labs cite most often since 2023. A benchmark counts as solved when the best model score reaches 85%. On 14.11, 109 of 173 are solved, or 63%. Until early 2025 the share stayed near 30%. In the last 12 months it grew by 20 points.

Benchmarks fall faster each year. In the first year after release, 17% of the 2023 benchmarks were solved, 30% of the 2024 benchmarks and 43% of the 2025 benchmarks. The median time to 85% fell from about 20 months to about 14.

In 2026, only 2 of 134 model launch posts named a benchmark from 2023, such as MMLU, HumanEval or GSM8K.

The author gives the most important caveat himself. Most "solutions" rest on claims by the labs. With only independent runs, SatBench is at 24%. Three of four first crossings past 85% come from a lab launch post, and independent runs later matched only 10 of 75.

Why it matters: a number in a model launch post tells less and less about the real difference between models.

A benchmark saturates fast, and the lab often gets the result in its own setup. To choose a model for a task, it is more useful to keep a small private test set and run it on each new version. The next item is an example.

TOPIC 6

Fable 5.2 against Opus 5.5 in Claude Code: the same result at 2.8 times the cost

SOURCES @garrytan, 14.11 @garrytan · g-harness-evals report, 12.11 GitHub

Garry Tan, the president of Y Combinator, keeps an open repository of tests for agent harnesses. The new run compares Fable 5.2, Fable 5.1 and Opus 5.5 in Claude Code 2.1.322. It uses 52 coding tasks with two attempts per task.

The authors wrote down the hypotheses and the decision rule before the run.

The result: Fable 5.2 solved 82.7% of the tasks, and Opus 5.5 solved 84.6%. The difference is -1.9 points with a range from -9.6 to +5.8, so the run shows no measurable difference. The cost per solved task differs: $1.83 for Fable 5.2 against $0.65 for Opus 5.5, or 2.8 times more. The report concludes that Opus 5.5 stays the default model for code.

The report shows its limits openly. The earlier run on Claude Code 2.1.312 did not know Fable 5.2, so it gave other numbers. This time the gap became smaller mostly because Opus scored lower. Opus already solves most tasks, so a clear difference needs harder tasks. With only 52 tasks, the smallest gap that the run can catch is 11.5 points.

Why it matters: a new flagship model is not always better for daily work. On typical tasks, the difference in quality is inside the noise, and the difference in cost is almost three times. The report also shows what a fair comparison looks like: rules written in advance, error ranges and a separate line about what the test cannot detect.

TOPIC 7

Google /teamwork: agents formalized the Weil conjectures in 1.8 million lines of Lean

SOURCES Google Research, 13.11 Research · HN, 14.11 Hacker News

Google described results of /teamwork, the multi-agent mode of its Antigravity environment on Gemini 4 models. The idea has two principles. First: many parallel attempts plus separate checkers, such as the Lean 4 proof checker, unchanged test suites and simulators. Second: each agent sees only the files it needs, and critics check the result from a clean context.

The biggest result is an autonomous, machine-checked formalization of the Weil conjectures, which Pierre Deligne proved in 1974. It has about 1.8 million lines of Lean and 28,800 subagent calls. The list also has the first formalization of the Birkhoff conjecture, open since 1927, and a proof of the Courtade-Kumar conjecture from information theory. In engineering, the agents rewrote the rest of the Zircon microkernel in safe Rust in two weeks.

This is about 800,000 lines that pass the kernel tests.

Separately, Google gave its review tool Deep Review to the authors of 578 papers for the STOC 2027 conference. In a survey of 70 authors, more than 30% say that the tool found substantive errors in proofs. All numbers come from Google itself.

Why it matters: all these results share one feature, formal verification. The model proposes, and Lean, tests or a simulator decide if the claim is true. This is the same scheme as in the story about the Gödel notes yesterday, but at the scale of millions of lines. Where no checker exists, such numbers are not possible yet.

TOPIC 8

Claude Code Projects: a coordinator gives work to parallel threads

SOURCES Anthropic, 13.11 Anthropic · Anthropic, 13.11 Anthropic · HN, 14.11 Hacker News

Anthropic redesigned projects in Claude Code. The user sets a goal and connects a repository. A coordinator divides the work into threads, and each thread is a separate cloud session of Claude Code on its own branch. Threads open PRs and run tests. Conflicts between them resolve as normal merge conflicts. A project has a shared memory and a library of files. The beta is open to part of the Pro and Max subscribers, and Anthropic promises a wider rollout within a week. Anthropic warns that several threads at once use up the limits faster.

On the same day, Claude Meetings came out in beta for the Team and Enterprise plans. Claude joins a Google Meet call as a visible participant, takes notes and makes documents during the talk. It keeps a transcript, not audio. If a person from outside the organization is on the call, Claude does not join by default.

Why it matters: in one week, Google /teamwork and Claude Code Projects came to the same scheme: a coordinator, parallel workers and a separate context for each worker. The scheme moves from research reports into products. The limit is also the same: parallel work multiplies cost, so it is useful to watch the cost per result.

TOPIC 9

Italy fined Meta €805 million for access to photos on Android

SOURCES BBC, 14.11 BBC · HN, 14.11 Hacker News

The Italian data protection authority fined Meta €805 million ($887 million). A Facebook feature suggested photos and videos from the phone gallery as ideas for posts. The app got gallery access through one general permission. The regulator found that about 20 million users did not know that this permission let Facebook analyze their photos. Meta has 60 days to turn the feature off by default and to ask for separate consent. The company will appeal.

Why it matters: the regulator judged the purpose for which Meta used the permission. In its view, one general permission for the gallery is not consent to AI analysis of photos. This rule applies to any app that adds AI features on top of permissions that users already gave.

TOPIC 10

US ground stop: L3Harris confirmed that it moves FAA communications from copper to fiber

SOURCES L3Harris, 14.11 L3harris · @garrytan, 14.11 @garrytan · HN, 15.11 Hacker News

Yesterday we wrote that on 13.11 the FAA stopped all departures in the US because of a "communication issue". Now L3Harris, the contractor that modernizes FAA communications, made a statement. The company writes that, with its telecom partners, it

moved more than 70% of the planned communications from copper wires to fiber. L3Harris helps the FAA with the investigation. The company does not name the cause of the stop, and no official links the stop to the conversion.

Why it matters: the statement does not say that the conversion caused the stop. But it shows that at the time of the failure, the system was partly old and partly new. Large migrations are often most fragile in the middle, when both systems work at the same time.

---

misc

- **Agents read the Internet Archive one line at a time:** Ethan Mollick asked Fable 5.2 to rebuild the 1941 poetry class of W. H. Auden. The agent found the right book in the Internet Archive and read it line by line with repeated full-text searches. Mollick writes that agent traffic will be large, and he donated to the archive.

  @emollick
- **Revolut:** the company published a postmortem of its five-hour global card outage. Gergely Orosz writes that there was no monitoring for database nodes that went down. 5 nodes in 3 regions gave no resilience when one node failed. @levelsio
- **Unsloth:** the company updated its compressed versions of Qwen3.8-27B to Dynamic V3.1. Unsloth says that accuracy is 5-10% better at the same size. The 4-bit model runs on 16-19 GB of VRAM or on a Mac with 24 GB.

  Unsloth · HN Hacker News
- **Webloc:** a Bloomberg investigation, which Citizen Lab cites, found new customers of the Webloc system for surveillance through ad data. Among them are the UK, the Netherlands, Germany, Mexico, Singapore, Saudi Arabia, the UAE and Israel. Citizenlab
- **HubSpot:** on 12.11, co-founder Brian Halligan came back as CEO, and the company calls this a "refounding".

  Dharmesh Shah writes daily notes about the process. @dharmesh
- **@amasad on the failures of the week:** "Anything that can go wrong will go wrong". Billions of interdependent processes make rare failures a daily event. @amasad
- **Monkey Island:** a site shows all 13 versions of the game side by side.

  Pages · HN Hacker News
