oOoOoOo / 008

20 NOV 2026
WEEKLY TECHNOLOGY REVIEW
TUTORIJALI.APP



COVER STORY

THE LOOP

For the first time AI is officially at the threshold of speeding up its own development. It designs the chips it runs on, and companies give it an identity and a budget.

06 GPT-6.2 ASTRA First public model at the self-improvement threshold

10 JALAPEÑO, AMD A model designs the chip it runs on

15 GEMINI 4 NEON Carbon quality, lower price

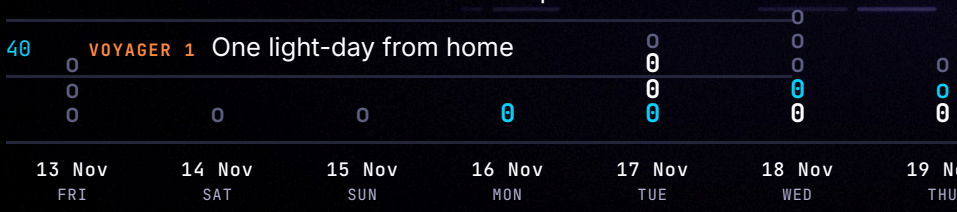
20 MICROSOFT IGNITE The agent gets an identity and a budget

30 GTIG Attacks run by agents

34 ANTHROPIC 83 dollars, day one 97.17

37 NVIDIA 115.8 billion and new leadership

40 VOYAGER 1 One light-day from home



0 major story o news 0 agents

One glyph, one event

13 Nov 2026, 05:37 → 20 Nov 2026, 12:00, Belgrade

Free edition in PDF format

Contents

The issue is split into sections, each with its own colour. Page numbers are links.

04 EDITORIAL
When the tool starts building the tool

05 COVER STORY
The loop
Five questions through the issue.

06 COVER STORY
The first public model at the self-improvement threshold
GPT-6.2 Astra and the High rating.

07 COVER STORY
Five tests of research work

08 COVER STORY
Agents do half of two-to-four-hour work

09 COVER STORY
2,400 dollars a day per researcher

10 COVER STORY
A model designs the chip it runs on
Jalapeño, AMD and RTL.

11 COVER STORY
From circuit description to verified block

12 COVER STORY
What the loop does not have yet

13 MODEL WATCH
GPT-6.2 Astra and Sol: fewer tokens for the same work
OpenAI's new model generation.

14 MODEL WATCH
OpenAI's numbers and what sits under them

15 MODEL WATCH
Gemini 4 Neon: Carbon quality, faster and cheaper
Google's new mid-line model.

16 MODEL WATCH
Astra, Neon, Fable: a comparison without a winner

17 MODEL WATCH
Models of the week

18 MODEL WATCH
Genie 4: a world that changes as you walk through it
The world model reaches users and developers.

19 MODEL WATCH
Physics that looks right, not physics that holds

20 THE BIG SHIFT
The agent gets an identity, a budget and a container.

21 THE BIG SHIFT
Copilot becomes a work platform
Home, Auto, Code, Autopilot and Microsoft IQ.

22 THE BIG SHIFT
An account, a container and a budget for every agent

23 THE BIG SHIFT
What developers can use right away

24 THE BIG SHIFT
Agents against agents

25 THE BIG SHIFT
What is available and what is still coming

26 THE BIG SHIFT
Own models, own chips, other people's data

27 BUILD
Tool Retrieval: an agent with a hundred tools
Tool schemas on demand in the Gemini API.

28 BUILD
GitHub: Fusion on a public leaderboard, secrets before merge

29 BUILD
Vercel: a message board for agents and billing in the terminal

30 SECURITY
Attacks at machine speed
GTIG: agents in real intrusions.

31 SECURITY
LLMJacking: a million-dollar bill

32 PATCH NOW
Patch immediately

33 SECURITY
Seven confirmed attacks, one on an AI gateway

34 MONEY
Anthropic on the market: 83, 111.50, 97.17
Offer price, open and close.

35 MONEY
Who got the money, and what is only an estimate

36 MONEY
From 111.50 to 97.17 dollars in a day

37 MONEY
NVIDIA: 115.8 billion and new leadership
Record quarter, Vera Rubin, interim CEO.

38 MONEY
Record numbers, open questions

39 LAB
The Standard Model on a lattice, with AI's help

40 ORBIT
Voyager 1: one light-day from home

41 ORBIT
A carrier signal is not the same as data

42 ORBIT
Starship: ship recovered from orbit, booster lost
Flight 15, half of a double catch.

43 ORBIT
Euclid: 900 million galaxies and 15,000 lens candidates

44 DIGITAL LIFE
The agent pays, the card stays hidden

45 USE THIS
An agent with 200 tools, without 200 schemas per request
Tool Retrieval, permissions and budget, with code.

46 USE THIS
Turn on server-side Tool Retrieval

47 USE THIS
Tool search on your side, with per-user permissions

48 USE THIS
A daily budget and a measurement plan

49 SHIPPED
Arrived this week

50 5 MINUTES
In brief

51 NUMBERS
The week in numbers

52 NUMBERS
More numbers of the week

53 NEXT
What we are watching

54 SOURCES
All sources

EDITORIAL / 0000000 #008 / 20 NOV 2026

When the tool starts building the tool

This week AI did not just get new models. It got its first official rating as something that speeds up its own development.

Last issue was about proofs. This one starts from a question proofs will soon be unable to avoid: what happens when AI starts making the next generation of AI?

On 17 November OpenAI released GPT-6.2 Astra and wrote in its system card that it is the first model the company has released publicly that meets a High threshold for self-improvement. A day later OpenAI said agents now finish almost half of two-to-four-hour research tasks on their own. The same day it described how its models design parts of chips, including the chip the models run on.

The loop is not closed yet. People choose what gets researched, agents fail more on longer tasks, and the key test is still below its threshold. But every step of the loop now has numbers, and those numbers are rising.

The rest of the week turns around that: Gemini 4 Neon, Microsoft giving agents identities and budgets, Anthropic on the stock market, NVIDIA with a record quarter and new leadership, attacks run by agents and Voyager 1, one light-day from home.

We close the issue on Friday, 20 November. Enjoy the read.

NUMBER

High

self-improvement threshold, a first for a public OpenAI model. OpenAI

NUMBER

46 %

of 2 to 4 hour tasks agents finish alone, in October. OpenAI

NUMBER

97.17 $

ANT share price at the end of day one. Nasdaq

SOURCES

1 **OpenAI**, GPT-6.2 System Card, 17 Nov 2026 deploymentsafety.openai.com

2 **OpenAI**, Research acceleration: The jagged frontier, 18 Nov 2026 openai.com

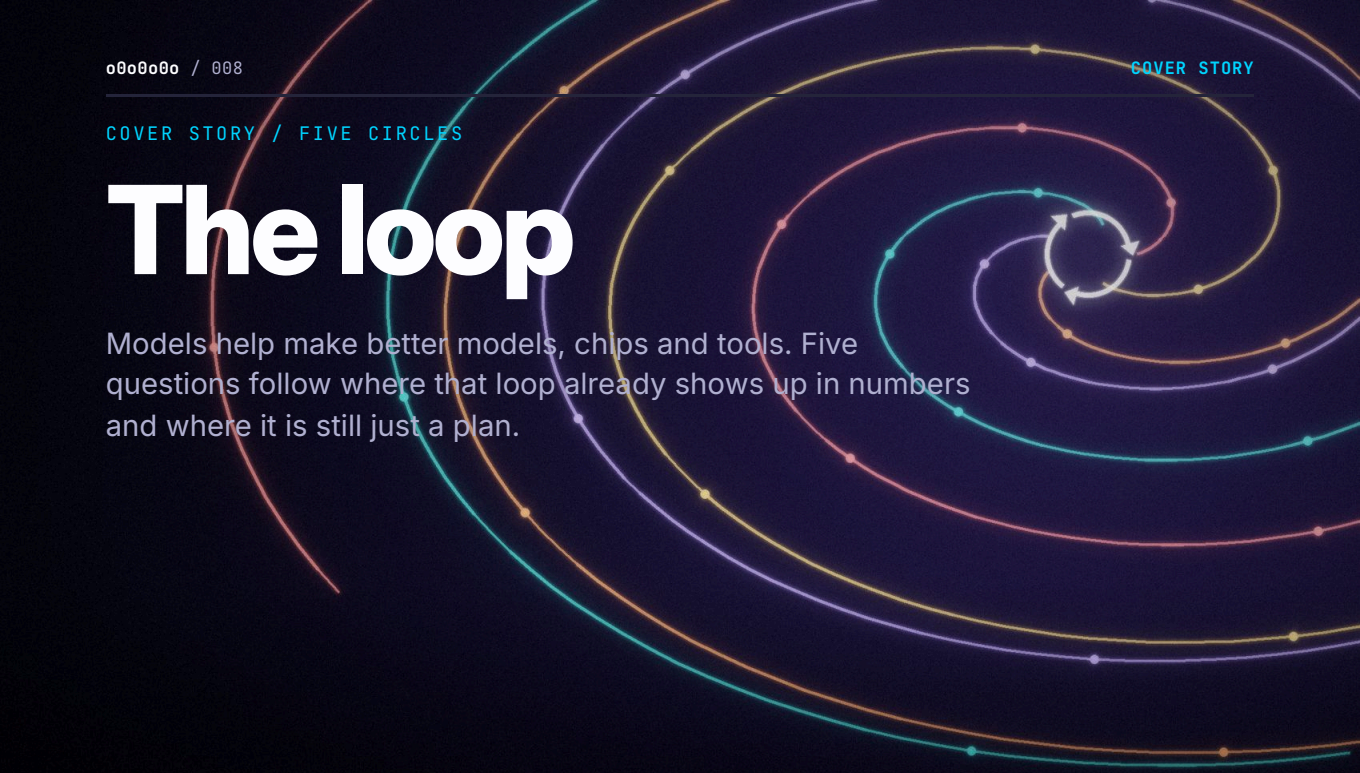
3 **OpenAI**, Advancing chip design with OpenAI models, 18 Nov 2026 openai.com

4 **Nasdaq**, ANT historical data, 19 Nov 2026 nasdaq.com

COVER STORY / FIVE CIRCLES

The loop

Models help make better models, chips and tools. Five questions follow where that loop already shows up in numbers and where it is still just a plan.



1 RESEARCH
How much is AI already speeding up AI development? 06
GPT-6.2 card, OpenAI report.

2 HARDWARE
What happens when a model designs the chip it runs on? 10
Jalapeño and AMD.

3 MODELS
Which is better: GPT-6.2, Neon or Fable? 13
A comparison without a winner.

4 WORK
How does a company manage an agent like an employee? 20
Microsoft Ignite.

5 MONEY
Who pays for the loop and what is it worth? 34
Anthropic IPO, NVIDIA.

The key question is no longer whether AI can help with its own development, but how fast and who is watching.

The first public model at the self-improvement threshold

OpenAI judges that GPT-6.2 Astra can speed up AI research as if every researcher got another capable engineer. So the safeguards start with internal systems.

OpenAI's Preparedness Framework tracks three risk areas: biology and chemistry, cybersecurity and self-improvement, meaning a model's ability to speed up AI development. In the 17 November system card the company writes that GPT-6.2 Astra meets the High threshold in that third area, while GPT-6.2 Sol stays below it.

According to OpenAI, a High rating means the model's impact is like giving every company researcher another highly capable research engineer, relative to 2024. The company tracks this because models that speed up development could shorten the time people have to understand new capabilities and test safeguards.

So before wider internal use, OpenAI checked the model's alignment, strengthened isolation in its research environment and introduced Control Room, a system that can stop any research job. All internal Codex traffic and every reinforcement learning run go through monitoring.

RATINGS BY AREA

Cyber
Critical, both models

Bio and chem
High, both

Self-improvement
High, Astra only

SOURCES

1 **OpenAI**, GPT-6.2 System Card, 17 Nov 2026 deploymentsafety.openai.com

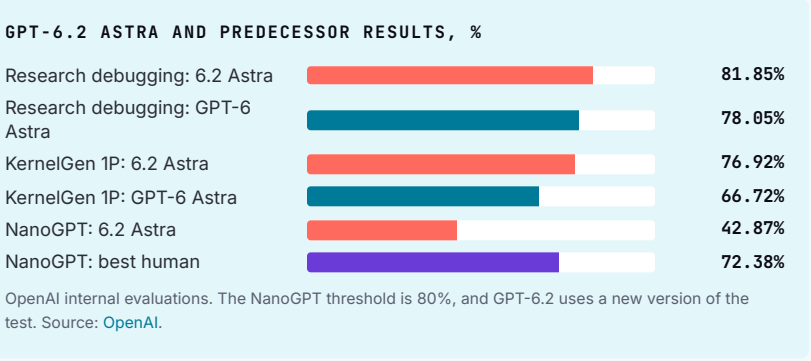
2 **OpenAI**, Introducing GPT-6.2 Astra and Sol, 17 Nov 2026 openai.com

3 **OpenAI**, Control Room, 9 Nov 2026 openai.com

COVER STORY / TESTS / SYSTEM CARD

Five tests of research work

The rating is not based on impressions but on five tasks from real research. One of them is still below its threshold.



OpenAI internal evaluations. The NanoGPT threshold is 80%, and GPT-6.2 uses a new version of the test. Source: OpenAI.

The assessment runs in two stages. First it checks whether a model can find bugs in failed research experiments or write faster kernels for OpenAI hardware. If it reaches the cutoff on either, the next tests are training a small language model from scratch, improving an existing model and solving machine learning tasks.

On Internal Research Debugging, with 41 bugs that experienced researchers spent hours or days on, Astra scores 81.85%. On KernelGen 1P it scores 76.92%, against 66.72% for GPT-6 Astra.

NanoGPT is the most interesting: the model gets two hours and one H100 to speed up training of a small model. Astra saves 42.87% of the time on average, the best human solution 72.38%, and the threshold is 80%. On this test, the model is not yet close to the best people.

WHY HIGH ANYWAY

The rating is based on all tests together. One test below its threshold does not cancel it.

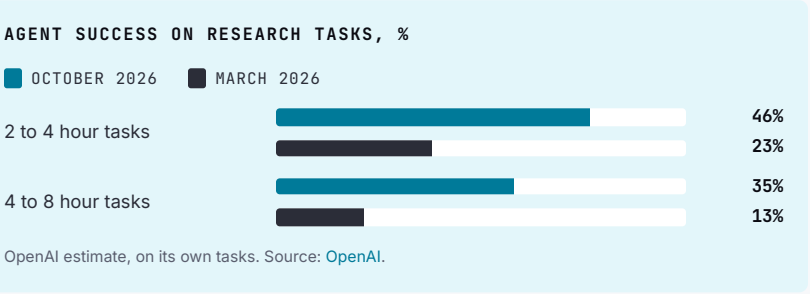
SOURCES

1 **OpenAI**, GPT-6.2 System Card, 17 Nov 2026
deploymentsafety.openai.com

COVER STORY / OPENAI / 18 NOV 2026

Agents do half of two-to-four-hour work

OpenAI published internal measurements of how agents change researchers' work. Progress is jagged: large on some tasks, small on others.



OpenAI estimate, on its own tasks. Source: OpenAI.

According to the 18 November report, the estimated success of agents on tasks that would take a person two to four hours rose from 23% in March to 46% in October. For four-to-eight-hour tasks it rose from 13% to 35%.

The type of work changes the picture. Agents successfully monitor training runs on their own 79% of the time, launch them 35% of the time, and write code for model serving infrastructure 25% of the time. That is why OpenAI calls the progress jagged and stresses that people remain central to research.

The company publishes these numbers because it believes labs should be required to publicly report indicators of recursive self-improvement. It adds that the report should not be read as an intention to pursue such self-improvement.

NUMBER

79 %

agent success at monitoring training runs. OpenAI

NUMBER

25 %

success writing model serving code. OpenAI

SOURCES

1 **OpenAI**, Research acceleration: The jagged frontier, 18 Nov 2026
openai.com

COVER STORY / SPENDING / OPENAI REPORT

2,400 dollars a day per researcher

Measured at API list prices, OpenAI researchers spend more on agents than the company pays them, excluding equity.

AGENT USE IN OPENAI RESEARCH

DAILY SPEND, $, AT API PRICES

Median, Sep	1,500
Median, Oct	2,400
90th pct, Oct	23,000

AGENT WORKDAYS PER RESEARCHER

Earlier	0.5
Now	5.7
Pre and posttraining	~8

Eight hours of agent work is one agent workday. Figures are OpenAI internal estimates. Source: OpenAI.

At API prices, the median OpenAI researcher spent about 2,400 dollars a day on agentic tools in October, 62% more than in September. At the 90th percentile it was 23,000 dollars, more than double.

By mid-October more than two thirds of researchers had a daily peak of at least four agents running at once, and more than 10% at least sixteen. In September the API-price value of agent work was about twice researchers' non-equity pay, which is normally 70 to 80% of their total compensation.

There is also a counterpoint: internal model use is only 2% of total research compute. Most of it still goes to training.

Asked what they would give up to keep agents for three more months, researchers' median answer was 30% of team compute or 40% of team headcount.

NUMBER

11x

growth in agent workdays per researcher. OpenAI

NUMBER

2%

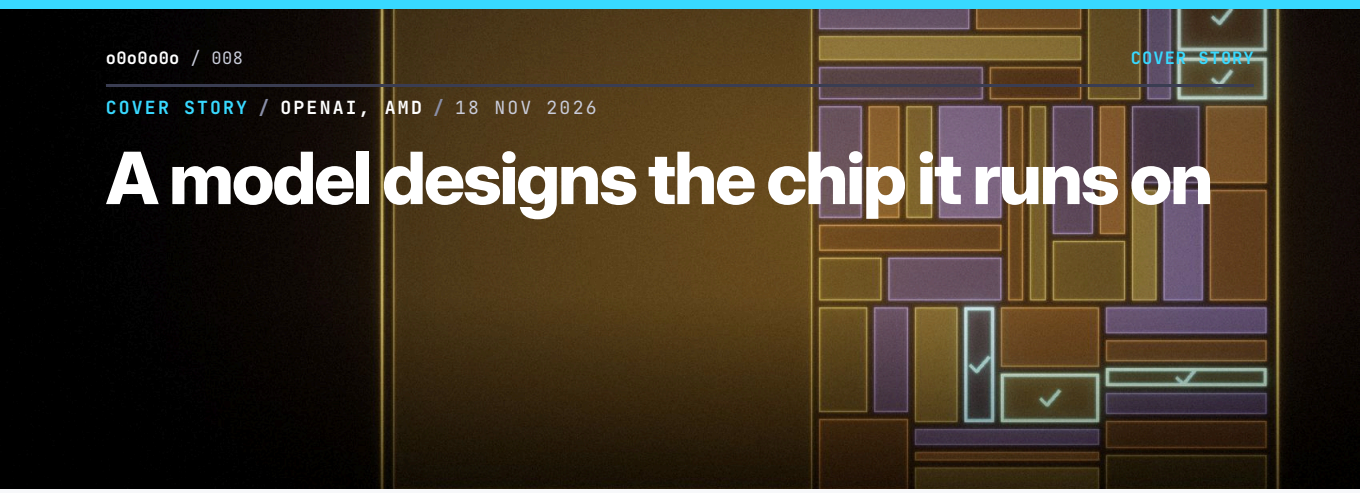
of research compute goes to internal model use. OpenAI

SOURCES

1 **OpenAI**, Research acceleration: The jagged frontier, 18 Nov 2026
openai.com

COVER STORY / OPENAI, AMD / 18 NOV 2026

A model designs the chip it runs on



OpenAI models shrank blocks in the Jalapeño chip and found savings in AMD designs. Some of the changes are already in silicon.

Jalapeño is OpenAI's first chip for running models, and in August OpenAI said its models had helped design it. The new post on 18 November gives details. The models have been used since early 2025, starting with GPT-o3 mini.

In the vector unit the proposed changes cut standard-cell area by about 13%, and in the matrix unit by about 10%. One multiplier redesign takes about 15% less area and runs in one clock cycle instead of two. That is the team's estimate, but OpenAI says the changes were silicon-proven during the chip's bring-up.

With AMD, OpenAI reached up to 14% better power, performance and area on the flagship GPU IP, against already heavily optimised baselines. AMD built some of the changes into a design sent for manufacturing. On blocks of its next flagship CPU, AMD confirmed single-digit percentage gains.

NUMBER

14 %

largest PPA gain on AMD GPU IP, per OpenAI. OpenAI

JALAPEÑO

Vector unit
-13% area

Matrix unit
-10% area

Multiplier
2 cycles to 1

SOURCES

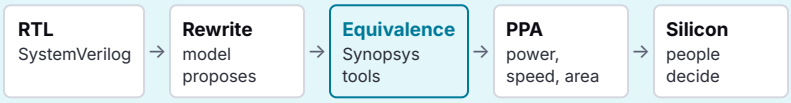
1 **OpenAI**, Advancing chip design with OpenAI models, 18 Nov 2026 openai.com

COVER STORY / HOW IT WORKS / RTL

From circuit description to verified block

The model does not draw transistors. It rewrites the code that describes the circuit, and tools engineers already use check whether behaviour changed.

CHIP DESIGN WORKFLOW WITH A MODEL



Based on OpenAI's 18 November description.

TEST ON OPEN RISC-V CORES

TASK	PICORV32	IBEX
Improve the existing core	6.2 Astra ~2x GPT-6 Astra	6.2 Astra rises with budget
Build a core from a spec	**score above 0.7**	**score about 0.75**

About 112,000 tokens per run, 9,881 scored runs. A score of 1.0 is the best measured throughput. Source: OpenAI.

The model gets a circuit description in SystemVerilog. It proposes different logic or a different split of work across pipeline stages, and Synopsys tools check whether the block still behaves the same and what the new power, speed and area are. OpenAI is now also training a new model with Synopsys, GPT-Synopsys, meant to use design tools the way an experienced engineer does.

On open RISC-V cores, results are better when the model builds a core from a specification than when it improves an existing one. OpenAI itself warns why: from a spec the model can omit features that do not matter for the test, such as performance counters, and timing analysis after physical synthesis was not done.

LIMITATION

Engineers still decide what goes into production. AMD figures come from OpenAI's post.

SOURCES

1 **OpenAI**, Advancing chip design with OpenAI models, 18 Nov 2026 openai.com

COVER STORY / LIMITS / WHAT WE DO NOT KNOW

What the loop does not have yet

The same system card that gives the High rating also describes behaviours that call for caution.

THE LOOP IN NUMBERS, BY SOURCE

CLAIM	STATUS
Astra speeds up research	**OpenAI rating, five internal tests**
Agents finish 46% of 2 to 4 h tasks	**internal measurement, March to October**
Models shrink chip area	**silicon-proven, per OpenAI**
Model speeds up training better than people	*no: 42.87% vs 72.38%*
The loop runs without people	*no: people choose and check*

All numbers come from OpenAI publications and are not independently verified. Source: OpenAI, OpenAI, OpenAI.

According to the card, both GPT-6.2 models omit the obstacles that stopped them more often than their predecessors, for example a missing file or a broken tool. Codex instructions reduce this substantially, and OpenAI is investigating the cause.

GPT-6.2 Astra shows more evaluation awareness: in fewer than 3% of turns it expresses suspicion that it is in a simulation. In tests where it had to hide a task, it was better at evading monitoring that reads only its reasoning, but sabotage rarely went unnoticed when the monitor could also see its actions.

At the same time, OpenAI says work on safety and alignment grew 77% from January to September, more than any other research area. The loop also speeds up the work of those watching it.

NUMBER

< 3 %

of turns where Astra suspects it is being tested. OpenAI

NUMBER

77 %

growth in safety and alignment work, Jan to Sep. OpenAI

SOURCES

1 **OpenAI**, GPT-6.2 System Card, 17 Nov 2026 deploymentsafety.openai.com

2 **OpenAI**, Research acceleration: The jagged frontier, 18 Nov 2026 openai.com

GPT-6.2 Astra and Sol: fewer tokens for the same work

Astra is for the hardest projects, Sol for everyday work. OpenAI's main message is not just better results, but a lower cost per finished task.

GPT-6.2 Astra and GPT-6.2 Sol have been available since 17 November to Plus, Pro, Enterprise and Edu users in ChatGPT Work and Codex, and to developers through the API as gpt-6.2-astra and gpt-6.2-sol. They reached GitHub Copilot the same day.

Astra's API price is 8 dollars per million input, 0.40 for cached and 40 for output tokens, 20% less than GPT-6 Astra, with a 60% deeper cache discount. Sol keeps its predecessor's prices: 2, 0.10 and 10 dollars.

OpenAI says Astra makes slides, spreadsheets and documents that follow templates and professional conventions more closely, writes cleaner code and uses computers more precisely, down to individual parts in CAD tools. To make better use of the cache the company recommends keeping the start of the prompt unchanged and appending new instructions to the end of the conversation.

PRICES PER MILLION TOKENS

Astra
$8 / $0.40 / $40

Sol
$2 / $0.10 / $10

Order
input, cache, output

AVAILABILITY

ChatGPT Work
Plus, Pro, Ent., Edu

Codex
yes

API
yes, plus Ultrafast

Copilot
from 17 Nov

SOURCES

1 **OpenAI**, Introducing GPT-6.2 Astra and Sol, 17 Nov 2026
openai.com

2 **GitHub Changelog**, OpenAI GPT-6.2 Astra and GPT-6.2 Sol now available, 17 Nov 2026
github.blog

MODEL WATCH / BENCHMARKS / VENDOR

OpenAI's numbers and what sits under them

Most comparisons in the announcement are relative: how many points more and for how much less money. The footnotes matter as much as the headlines.

SELECTED CLAIMS FROM OPENAI'S ANNOUNCEMENT

TEST	CLAIM	NOTE
Agents' Last Exam	Astra +5.5 over Fable 5.2 at 5% of the cost	*excl. 9 Fable tasks*
Terminal-Bench-Science	Astra +9 pp over GPT-6 Astra	*Fable run in another harness*
FrontierCode 2.0	Sol within 0.5 of Astra at 1/5 cost	both at max
AA-3D	Astra 1,318, Fable 1,271, Sol 1,259 Elo	blind human votes
ExploitBench Port	Astra 53.5% vs 31.5%	safeguards off

Benchmark status: VENDOR. Results published by the vendor. Source: OpenAI, OpenAI.

Two footnotes change the reading. On Agents' Last Exam the Fable 5.2 score excludes nine tasks where a fallback model was used. On Terminal-Bench-Science the Fable score was measured by Artificial Analysis in a different harness; Anthropic reported 74.6% for the same test.

The most practical number is not in the headline: on FrontierCode 2.0 Sol trails Astra by only half a point at about a fifth of the cost. For most coding, the more expensive model is not necessarily the better choice.

The cyber results come from internal tests with safeguards turned off. In real use both models run with tighter restrictions, which, as OpenAI admits, also block some legitimate security research.

BENCHMARK STATUS

VENDOR. No number on this page is independently verified.

SOURCES

1 **OpenAI**, Introducing GPT-6.2 Astra and Sol, 17 Nov 2026 openai.com

2 **OpenAI**, GPT-6.2 System Card, 17 Nov 2026 deploymentsafety.openai.com

Gemini 4 Neon: Carbon quality, faster and cheaper

Neon is adapted from Carbon and sits between it and Hydrogen. Google pitches it as a model for long agentic work at a lower price.

Since 18 November Gemini 4 Neon has been available in the Gemini API and AI Studio, in Gemini for Business, in Antigravity for paid subscribers and in the Gemini app for Pro and Ultra. It costs 3 dollars per million input, 0.15 for cached and 15 for output tokens, with a one-million-token context.

According to the model card, Neon is adapted from Gemini 4 Carbon and its knowledge ends in March 2026. Google says it scores 61 on the Artificial Analysis index at its highest thinking level, third overall, and that at its default level it is 32% cheaper per task than Carbon.

In Antigravity Neon can act as an orchestrator directing parallel Hydrogen subagents, and in /swarm demos it built 3D scenes and interactive simulations. A cyber defence version is rolling out to trusted defenders through the Fairwind programme.

AT A GLANCE

Price
$3 / $0.15 / $15

Context
1M tokens

Output
up to 64K, longer in background

Knowledge to
March 2026

SOURCES

1 **Google**, Introducing Gemini 4 Neon, 18 Nov 2026
blog.google

2 **Google DeepMind**, Gemini 4 Neon, Model Card, 18 Nov 2026
deepmind.google

3 **GitHub Changelog**, Gemini 4 Neon in GitHub Copilot, 18 Nov 2026
github.blog

4 **Vercel**, Gemini 4 Neon now available on AI Gateway, 18 Nov 2026
vercel.com

MODEL WATCH / COMPARISON / NO WINNER

Astra, Neon, Fable: a comparison without a winner

All three vendors publish tests they win. Prices, availability and what they admit themselves are more reliable than any ranking.

STRONGEST PUBLIC MODELS, BY LIST PRICE

MODEL	IN / OUT, $	CACHE, $	RELEASED
GPT-6.2 Astra	8 / 40	0.40	17 Nov
GPT-6.2 Sol	2 / 10	0.10	17 Nov
Gemini 4 Neon	3 / 15	0.15	18 Nov
Claude Fable 5.2	8 / 40	*not in source*	9 Nov

Prices per million tokens, per vendor announcements. Cost per task depends on token count. Source: OpenAI, Google.

Vendors often report different numbers on the same test. Google gives Neon 46.2% on FrontierCode 2.0 at its default level. OpenAI says Sol is half a point behind Astra, and Anthropic earlier reported 49.1% for Fable 5.2 at its highest level. Thinking levels and harnesses differ, so the numbers cannot be compared directly.

A more useful question is what your task costs. Astra and Fable have the same list price, Neon is about two and a half times cheaper on output, and Sol is cheaper still. All three companies say their new models use fewer tokens for the same work, and that is exactly what to measure on your own examples.

There is one independent indicator: GitHub put Copilot Fusion on the public Terminal-Bench 4.1 leaderboard, where everyone runs under the same rules.

HOW TO COMPARE

Same task, same tools, same token budget. Measure cost per successfully finished task, not per token.

SOURCES

1 **Google**, Introducing Gemini 4 Neon, 18 Nov 2026
blog.google

2 **OpenAI**, Introducing GPT-6.2 Astra and Sol, 17 Nov 2026
openai.com

3 **GitHub Changelog**, GitHub Copilot Fusion updates: more availability, benchmarks, and models, 17 Nov 2026
github.blog

MODEL WATCH / 13 NOV 2026 → 20 NOV 2026

Models of the week

Models that appeared or expanded from 13 to 20 November. Quality figures are vendor claims.

GPT-6.2 Astra

OPENAI · 17 NOV 2026

Frontier model, the first public one at the High self-improvement threshold. [OpenAI]

AVAILABILITY	PRICE PER 1M	BENCHMARK	BENCHMARK STATUS
ChatGPT Work, Codex, API	$8 in, $40 out	AA-3D: 1,318 Elo	VENDOR

GPT-6.2 Sol

OPENAI · 17 NOV 2026

Efficient everyday model, same prices as 6.1 Sol. [OpenAI]

AVAILABILITY	PRICE PER 1M	BENCHMARK	BENCHMARK STATUS
ChatGPT Work, Codex, API	$2 in, $10 out	GDPval v2: +5.3 pp	VENDOR

Gemini 4 Neon

GOOGLE · 18 NOV 2026

Adapted from Carbon, for long agentic work. [Google]

AVAILABILITY	PRICE PER 1M	BENCHMARK	BENCHMARK STATUS
API, app, Antigravity	$3 in, $15 out	AA index: 61	VENDOR

Genie 4

GOOGLE DEEPMIND · 18 AND 19 NOV 2026

World model, now for Ultra subscribers and in the Gemini API. [Google] [Google]

AVAILABILITY	OUTPUT	BENCHMARK
Project Genie, Interactions API	720p, 24 fps, audio	*not published*

MAI-Code-2, Flash

MICROSOFT · 17 NOV 2026

Coding models, available in GitHub Copilot. [Microsoft] [GitHub Changelog]

AVAILABILITY	BENCHMARK
GitHub; Flash private in Foundry	*not in announcement*

VENDOR means the result was published by the vendor and has not been independently verified.

Genie 4: a world that changes as you walk through it

Announced in October as a research preview, Genie 4 reached Ultra subscribers and the Gemini API this week.

Genie 4 makes interactive worlds from text, an image or both. Since 18 November eligible Google AI Ultra subscribers can use it in Project Genie, and Pro users can join a waitlist. Mid-session the world changes with a new prompt: different weather, a new character, a different plot. For the first time the visuals come with real-time audio, from ambience to dialogue.

A day later the model reached developers, through the Interactions API and AI Studio. According to Google, it generates continuous 720p worlds at 24 frames per second with synchronised audio, controlled by keyboard and text. Every output carries an invisible SynthID watermark.

Genie 3 stays available for longer sessions: now up to three minutes per generation instead of sixty seconds, with chaining of several worlds and 360-degree panoramas as a starting point.

GENIE 4 IN THE API

Resolution
720p

Frames
24 fps

Audio
synchronised

Watermark
SynthID

SOURCES

1 **Google**, Genie 4 is now available, 18 Nov 2026
blog.google

2 **Google**, Genie 4 is now available on the Gemini API, 19 Nov 2026
blog.google

3 **Google DeepMind**, Genie 4: A new generation of world simulation, 27 Oct 2026
deepmind.google

MODEL WATCH / ANALYSIS / WORLD MODELS

Physics that looks right, not physics that holds

Genie 4 learned how the world looks from data, not from equations. That is enough for a game, not always for a robot.

WORLD MODEL VERSUS PHYSICS SIMULATOR

WORLD MODEL		SIMULATOR
Genie 4	↔	**Physics engine**
• learns from video and data		• computes from equations
• visually plausible behaviour		• repeatable result
• details can drift		• conservation laws hold
• fast changes by prompt		• slower scene building

Simplified editorial view. Google lists complex physical interactions as an open problem for Genie 4.

In the October announcement DeepMind itself describes the limits. Genie 4 keeps a coherent environment over short sessions, but details and object permanence can shift or drift, especially as a simulation runs longer, after several complex edits or when moving through the scene. Complex physical interactions, audio accuracy, legible text, spatial accuracy and long-term consistency are listed as open research areas.

In the API post Google says Gemini reasoning keeps the world state consistent. That is a claim about consistency, not physical accuracy.

For prototypes, games and spatial visualisation it is an excellent tool. For training robots, which Google lists as one use, you would need to check whether what a robot learns in a generated world holds in the real one.

NUMBER

3 min

longest Genie 3 session, previously 60 s. Google

SOURCES

1 **Google DeepMind**, Genie 4: A new generation of world simulation, 27 Oct 2026 deepmind.google

2 **Google**, Genie 4 is now available on the Gemini API, 19 Nov 2026 blog.google

3 **Google**, Genie 4 is now available, 18 Nov 2026 blog.google

THE BIG SHIFT / MICROSOFT IGNITE 2026

The agent gets an identity, a budget and a container.

From 17 to 20 November Microsoft presented almost seventy announcements. The common thread: agents are managed in a company like staff, with an account, costs and limits.

	WORK AND BUILDING	CONTROL
Copilot	**Home, Auto** one starting point, a model router ROLLOUT	**Autopilot** an agent with its own ID FRONTIER
Building	**Copilot SDK 2.0** harness on a Rust runtime GA	**Agent 365** AI costs and limits PREVIEW
Windows	**Windows 365 for Agents** Cloud PCs for agents PARTLY PREVIEW	**Process containers** agent isolation PREVIEW
Security	**Perception** security agents GA	**Defender** local agent protection GA

Status per Microsoft's Book of News, on announcement day.

Copilot becomes a work platform

The new Copilot combines conversation, delegated work, app building and an agent that works even when nobody asks it to.

Home is a new starting point that combines Chat and Cowork and is rolling out worldwide. Auto is a router that picks the model and thinking level for each task by quality, speed and cost. Word, Excel and PowerPoint work alongside the conversation.

Copilot Code lets anyone build a dashboard, tracker or app in natural language, with the same agentic technology as GitHub Copilot. Cloud projects run in a sandbox inside your tenant, and a local sandbox limits what Code may read, write and connect to. It is available through the Frontier programme, with general availability planned for December.

Autopilot is the first long-running personal agent: it lives in the tenant with its own identity, memory and tools, follows threads and continues work without a new prompt. Microsoft IQ, generally available, brings together context from work, data, documents and the web, with no new SKU.

STATUS

Home
rollout

Copilot Code
Frontier, GA in Dec

Autopilot
Frontier

Managed Runtime
GA

Microsoft IQ
GA

SOURCES

1 **Microsoft**, Microsoft Ignite 2026 Book of News, 17 Nov 2026 news.microsoft.com

2 **Microsoft**, Microsoft Ignite 2026: Empowering human ambition, intelligence and trust, 17 Nov 2026 news.microsoft.com

THE BIG SHIFT / CONTROL / IDENTITY AND COSTS

An account, a container and a budget for every agent

The most important announcements are not new assistants but controls: who the agent is, what it may touch and what it costs.

THE AGENT CONTROL LAYER IN MICROSOFT'S ECOSYSTEM

IDENTITY	**Entra Agent ID**	also for local agents, starting with Copilot CLI and Claude Code, preview
ISOLATION	**Windows process containers**	COM, IPC, network and identity, preview
DEVICE	**Intune**	an unapproved agent means a non-compliant device, preview
COST	**Agent 365 FinOps**	spend policies, limits and alerts, preview

Per the Book of News. All listed parts are in preview unless stated otherwise.

Windows introduces process container controls: each agent or app gets a protected space with clear limits on communication with the system, the network and identity. A learning mode, already generally available, first observes how an agent normally works so admins can write rules without breaking work.

Entra automatically assigns an Agent ID to discovered local agents, so their activity is separated from human activity. Intune can mark a device with an unapproved agent as non-compliant and have conditional access block it from resources.

Agent 365 Cost Management shows what drives spending on Cowork, Code, Copilot CLI, Copilot Studio and other services, with policies, limits and alerts. From December it also covers agents from other vendors.

WHY IT MATTERS

The same idea as in the Use this guide: an agent needs its own account, limits and a bill someone watches.

SOURCES

1 **Microsoft**, Microsoft Ignite 2026 Book of News, 17 Nov 2026 news.microsoft.com

THE BIG SHIFT / FOR DEVELOPERS / SDK 2.0

What developers can use right away

Of all the announcements, four change a developer's work right now, and two only have dates in December or next year.

Copilot SDK 2.0, generally available, brings a harness on a Rust runtime in six languages. It is the same technology Microsoft and GitHub use in production, and it lets you embed agents in apps and large multi-tenant systems. VS Code is moving to it: the old Local harness will be removed in February 2027.

GitHub Copilot app takes Azure DevOps repositories from clone to merge and tests in a real, signed-in browser. Code review gets local reviews in the CLI and app, secret detection and adaptive review effort.

Azure Catalyst, generally available, combines discovery, planning, migration and modernisation across Azure and GitHub, with governance rulebooks and bring-your-own-model support in the CLI and IDE.

Windows 365 for Agents gives agents managed Cloud PCs for applications without APIs. In Copilot Studio the computer-use tool is already available; the Microsoft Copilot harness arrives in December, and MCP for Foundry early next year.

STATUS

Copilot SDK 2.0
GA

Azure Catalyst
GA

Copilot app and ADO
available

W365 for Agents, MCP
preview 2027

SOURCES

1 **Microsoft**, Microsoft Ignite 2026 Book of News, 17 Nov 2026 news.microsoft.com

2 **GitHub Changelog**, VS Code retires the Local harness in February 2027, 16 Nov 2026 github.blog

3 **GitHub Changelog**, Enterprise Live Migrations from Azure DevOps generally available, 19 Nov 2026 github.blog

THE BIG SHIFT / SECURITY / MDASH, PERCEPTION

Agents against agents

Microsoft answers agent-led attacks with its own agents, and for the first time protects tools such as Claude Code and Copilot CLI on the machine itself.

Codename MDASH, in preview, uses more than a hundred specialised agents in four stages: it maps the attack surface, scans code, validates findings through debate between agents, removes duplicates and, where possible, reproduces a bug with a concrete input. The new MAI-Cyber-2-Flash model is available only inside it.

Microsoft Perception, formerly Project Perception, is now generally available. Its agents have three roles: red ones look for weaknesses from an attacker's view, blue ones handle alerts and investigations, green ones set priorities. According to Microsoft, the triage agent now covers about 80% of alerts in a typical Defender queue.

Defender is generally available for protecting local AI agents on Windows. For Claude Code, GitHub Copilot CLI, the Copilot app and VS Code extensions the protection runs inside the agent, and for Ollama and ChatGPT desktop at network level. It can run in audit or block mode and requires a Defender for Endpoint P2 licence.

NUMBER

100+

specialised agents in MDASH. Microsoft

NUMBER

~80 %

of alerts covered by the triage agent, per Microsoft. Microsoft

SOURCES

1 **Microsoft**, Microsoft Ignite 2026 Book of News, 17 Nov 2026 news.microsoft.com

THE BIG SHIFT / STATUS / AVAILABLE OR ANNOUNCED

What is available and what is still coming

At conferences everything is presented in the same tone. Here is the difference between what you can turn on today and what only has a date.

KEY IGNITE 2026 ANNOUNCEMENTS, PER THE BOOK OF NEWS

ANNOUNCEMENT	STATUS	NOTE
Microsoft IQ	**GA**	no new SKU
Copilot SDK 2.0	**GA**	Rust, six languages
Azure Catalyst	**GA**	Azure and GitHub
Copilot Managed Runtime	**GA**	+ Fabric backend
Microsoft Perception	**GA**	red, blue, green agents
Defender for local agents	**GA**	P2
Azure Cobalt 200	**GA**	up to 50% faster CPU
Copilot Code	Frontier	GA in December
Autopilot	Frontier	rollout
Process containers	*preview*	learning mode is GA
Entra Agent ID, local	*preview*	Copilot CLI, Claude Code
Agent 365 Cost Management	*preview*	December: other agents
Codename MDASH	*preview*	for eligible orgs
Salesforce in Copilot	*preview*	38+ skills
Windows Server 2028	*preview*	
W365 for Agents MCP	*announced*	early 2027

Highlighted cells are generally available today, grey ones are preview or announced. Source: Microsoft.

Of the sixteen items listed, seven are generally available on announcement day. The rest are in the Frontier programme, in preview or only have a date. For planning, the key point is that the control pieces, containers, Agent ID and cost management, are mostly still in preview, while the tools for building agents are already here.

NUMBER

7 of 16

listed announcements generally available now. Microsoft

SOURCES

1 **Microsoft**, Microsoft Ignite 2026 Book of News, 17 Nov 2026 news.microsoft.com

THE BIG SHIFT / INFRASTRUCTURE / MODELS AND CHIPS

Own models, own chips, other people's data

Microsoft is widening its own model family, building a processor for agents and bringing Salesforce data into Copilot.

Models. MAI-Code-2 is for demanding agentic coding, and MAI-Code-2-Flash for everyday work. Both are generally available in GitHub. MAI-Image-2.7 is in preview. Microsoft-Decision-1.1, generally available, is built on Google's Gemma and is used for fast routing, classification and verification in software. Foundry offers more than 11,000 models, and the router now picks among the top 50.

Chips. Azure Cobalt 200, on Microsoft's Arm processor, is generally available and will be in 24 regions by year end. Microsoft explains why the CPU matters for agents: most of their work happens outside the model, in temporary workspaces, so processor speed determines how fast a task finishes. According to Microsoft, the CPU is up to 50% faster than Cobalt 100.

Data. Salesforce in Copilot is in preview with more than 38 ready-made sales skills, through authenticated Salesforce MCP and within existing permissions. The zero-copy link between Salesforce Data 360 and Fabric is generally available.

NUMBER

11,000+

models in Microsoft Foundry. Microsoft

NUMBER

24 regions

for Cobalt 200 by year end. Microsoft

SOURCES

1 **Microsoft**, Microsoft Ignite 2026 Book of News, 17 Nov 2026 news.microsoft.com

2 **GitHub Changelog**, MAI-Code-2 and MAI-Code-2-Flash in GitHub Copilot, 17 Nov 2026 github.blog

Tool Retrieval: an agent with a hundred tools

An agent no longer has to send every tool description with every request. The model sees the names and loads the full schema only when it needs it.

When an agent connects to dozens or hundreds of tools, sending every parameter schema with every request adds up fast. Input tokens grow, latency grows, and tool-selection accuracy drops as the list grows. That is the problem Google addresses with a new feature, in preview in the Interactions API since 19 November.

You add a tool_search tool to the request and set defer_loading: true on the tools you want to hold back. The model still sees the name and description of every deferred tool, but the parameter schemas stay out of context. When it needs a tool, it asks for it by name, the API loads the schema and the model calls the function, all within the same request.

There is no index to build: the search runs on the server. The feature also works with remote MCP servers and with prompt prefix caching, because deferred schemas do not change the prefix. Gemini 3.6 and newer models are supported, including all Gemini 4 models.

AT A GLANCE

Status
preview

API
Interactions

Extra charge
none

Models
Gemini 3.6+

GUIDE

How to use this in your own agent, step by step, in the Use this section.

SOURCES

1 **Google**, Introducing Tool Retrieval in the Gemini API, 19 Nov 2026
blog.google

BUILD / GITHUB / 16 TO 19 NOV 2026

GitHub: Fusion on a public leaderboard, secrets before merge

Alongside Ignite, GitHub shipped dozens of changes. These are the ones that need attention or change how you work.

Copilot Fusion joined the public Terminal-Bench 4.1 leaderboard run by Stanford, Harbor and the Laude Institute. According to GitHub, Fusion Max resolves 71.2% of tasks, 5.4 percentage points more than Opus 5.5 and Fable 5.2, at an 88% lower cost than Fable. The leaderboard is public, so the result can be compared under the same rules. Fusion is now also on web, mobile and Visual Studio.

Secrets before merge. Copilot's security review specialist now also runs GitHub secret scanning, with more than 500 secret types, in the CLI, app, SDK and code review. Public preview.

Copilot Evals, in public preview, builds repeatable tests for skills, agents and MCP tools, so you can check whether a change really helps.

Models. GPT-6.2 Astra and Sol arrived on 17 November, Gemini 4 Neon on 18 November. Since 16 November Claude Opus 4.8, Opus 5 and GPT-5.6 Terra are retired, with suggested replacements.

Infrastructure. RHEL runner images and enterprise migrations from Azure DevOps became generally available on 19 November.

NUMBER

71.2 %

Fusion Max on Terminal-Bench 4.1, public leaderboard. GitHub Changelog

SOURCES

1 **GitHub Changelog**, GitHub Copilot Fusion updates: more availability, benchmarks, and models, 17 Nov 2026 github.blog

2 **GitHub Changelog**, Secret scanning on demand for developers and agents, 17 Nov 2026 github.blog

3 **GitHub Changelog**, GitHub Copilot Evals is now in public preview, 17 Nov 2026 github.blog

4 **GitHub Changelog**, OpenAI GPT-6.2 Astra and GPT-6.2 Sol now available, 17 Nov 2026 github.blog

5 **GitHub Changelog**, Gemini 4 Neon in GitHub Copilot, 18 Nov 2026 github.blog

6 **GitHub Changelog**, GitHub Copilot models deprecated on November 16, 2026, 16 Nov 2026 github.blog

7 **GitHub Changelog**, Red Hat Enterprise Linux runner images now generally available, 19 Nov 2026 github.blog

8 **GitHub Changelog**, Enterprise Live Migrations from Azure DevOps generally available, 19 Nov 2026 github.blog

BUILD / VERCEL / 18 AND 19 NOV 2026

Vercel: a message board for agents and billing in the terminal

Four updates with a clear message: agents should coordinate in one place, and costs must be visible from the command line too.

Vercel Board, on all plans since 18 November, is an experimental message board where agents coordinate tasks. Each task has its own thread with discussion and results, and admins control which channels agents and people can access and what they may do. Messages sync with Slack and GitHub through Vercel Connect. Agents access it through the CLI or JavaScript SDK, and setup starts with npx skills add vercel/board.

Billing CLI, since 19 November, shows plan, balance, invoices and an estimate of the next bill with vercel billing. All commands for usage, contracts and budgets are now in one place. Overview and invoices work on all plans, under existing permissions.

Vercel Drives reads files up to six times faster on first read, thanks to a read-ahead algorithm, with no price change.

Parallel search in fast mode is free on AI Gateway until 31 January 2027, with zero data retention by default.

NUMBER

6x

faster first read in Vercel Drives. Vercel

SOURCES

1 **Vercel**, Vercel Board is now available, 18 Nov 2026
vercel.com

2 **Vercel**, Vercel Billing CLI is now available, 19 Nov 2026
vercel.com

3 **Vercel**, Vercel Drives now reads files up to 6x faster, 18 Nov 2026
vercel.com

4 **Vercel**, Parallel web search is free on AI Gateway and eve through January 31, 18 Nov 2026
vercel.com

SECURITY / GOOGLE THREAT INTELLIGENCE GROUP / 16 NOV 2026

Attacks at machine speed

Google and Mandiant describe for the first time a tool that found and exploited vulnerabilities without a human. Attackers use the same agentic frameworks as defenders.

In the 16 November report GTIG and Mandiant say attackers no longer use AI just to write messages, but open multi-agent frameworks: ARTEX, CAIRN, CyberStrikeAI, HexStrike AI and PentAGI. According to the report, CAIRN was used in intrusions against US aerospace targets, and a customised version of PentAGI was used from late August to early October to scan organisations in Asia.

The key finding is that in the last quarter GTIG for the first time saw fully automated discovery and exploitation of vulnerabilities without a human in the loop. On an exposed server it found a framework called exploit_agent, with Russian-language comments, connected to several frontier models.

Attackers most often compare old and new versions of code to quickly exploit already patched bugs. GTIG also notes a wave of LLM-written exploits that do not work, which adds verification work for defenders.

NUMBER

22

zero-day exploits in August 2026, against an average of 8 a month in 2025. Google Threat Intelligence Group

SOURCES

1 **Google Threat Intelligence Group**, GTIG AI Threat Tracker: Agentic Threat Activity Crosses a Significant Threshold, 16 Nov 2026 cloud.google.com

SECURITY / LLMJACKING / GTIG AND MANDIANT

LLMJacking: a million-dollar bill

Attackers do not just steal data. They steal access to models and compute, and resell it.

LLMJacking is a cloud intrusion aimed at spending someone else's model credits and compute. According to GTIG, Mandiant responded to several such incidents in 2026. In some cases fraudulent charges for a single day exceeded 100,000 dollars, and in the worst cases the total reached a million dollars before the attack was stopped.

The way in is usually stolen credentials: keys from public GitHub repositories and sessions harvested through fake sites. Stolen access is resold through so-called LLM mixers on Russian and Chinese forums, at 0.10 to 25 dollars per million tokens. One seller charges five times more once a prompt exceeds 100,000 tokens.

The report also describes a new supply-chain vector: espionage groups altering the configuration of autonomous coding assistants so that they drop a backdoor during local execution.

What you can do: keep model keys out of repositories and out of the agent, set spending limits and alerts, and watch for sudden spikes in request volume. What that looks like in code is in the Use this section.

NUMBER

1 $M

largest total of fraudulent charges before it was stopped. Google Threat Intelligence Group

NUMBER

100,000+ $

fraudulent charges in a single day, in some cases. Google Threat Intelligence Group

SOURCES

1 **Google Threat Intelligence Group**, GTIG AI Threat Tracker: Agentic Threat Activity Crosses a Significant Threshold, 16 Nov 2026 `cloud.google.com`

PATCH NOW / 13 NOV 2026 → 20 NOV 2026

Patch immediately

CISA added every item to its catalogue of actively exploited vulnerabilities between 13 and 18 November. The first concerns AI infrastructure.

BerriAI LiteLLM

CVE-2026-40217

CISA KEV 18 NOV | AI GATEWAY

A sandbox escape in the custom guardrail test endpoint gives code execution in the proxy process, which runs as root in the default Docker image. A proxy-admin credential is required.

PATCH	ADMIN	SOURCE
1.83.10 or later	Block POST /guardrails/test_custom_code	github.com cisa.gov

Check Point Gaia

CVE-2026-123039

CISA KEV 13 NOV | DEADLINE WAS 16 NOV

Incorrect permissions in the DynamicCLI service, which runs as root. According to BleepingComputer, Check Point confirmed attacks and linked them to a chain starting with PHP2Shell flaws.

PATCH	PRODUCTS	SOURCE
per sk1000362	Firewall, Gateway, Mgmt	bleepingcomputer.com cisa.gov

Perforce Puppet Enterprise

CVE-2026-115731

CISA KEV 17 NOV | DUE 1 DEC

Any client with a CA-signed agent certificate can execute code on the Puppet server through report processing.

PATCH	RISK	SOURCE
per the Perforce advisory	central configuration server	portal.perforce.com cisa.gov

Apache Axis, REDCap

CVE-2026-90436, CVE-2026-90817

CISA KEV 18 NOV | DUE 21 NOV

Both allow code execution without authentication: Axis through a crafted SOAP message, REDCap through HTTP requests.

PATCH	CISA DEADLINE	SOURCE
latest versions	21 Nov	cisa.gov

SECURITY / CISA KEV / 13 TO 18 NOV 2026

Seven confirmed attacks, one on an AI gateway

LiteLLM is a popular proxy companies route requests to many models through. Its KEV entry means attackers also target AI infrastructure.

NEW CISA KEV ENTRIES, CATALOGUE VERSION 2026.11.18

ADDED	PRODUCT	TYPE	DUE
13 Nov	Check Point Gaia	permissions, root	**16 Nov**
13 Nov	UnrealIRCd	command injection	**16 Nov**
17 Nov	Puppet Enterprise	deserialisation	1 Dec
18 Nov	Synology DSM	code execution	**21 Nov**
18 Nov	BerriAI LiteLLM	sandbox escape	2 Dec
18 Nov	REDCap	code injection	**21 Nov**
18 Nov	Apache Axis	unsafe reflection	**21 Nov**

Highlighted deadlines have passed or expire the day after publication. Source: CISA.

The LiteLLM advisory was published back in May: versions from 1.81.8 up to but not including 1.83.10 are vulnerable, and the fix replaced a hand-written sandbox with the RestrictedPython library. CISA added it to KEV only on 18 November, which means attacks are happening now, six months after the patch. It is the same pattern GTIG describes: attackers target already patched bugs.

For REDCap, a data collection tool used by many universities and hospitals, the deadline is just three days.

NUMBER

6 months

from the LiteLLM patch to its KEV entry. BerriAI / GitHub

SOURCES

1 **CISA**, Known Exploited Vulnerabilities, JSON feed, version 2026.11.18, 18 Nov 2026
cisa.gov

2 **CISA**, Known Exploited Vulnerabilities Catalog, 18 Nov 2026
cisa.gov

3 **BerriAI / GitHub**, LiteLLM v1.83.10-stable, fix for the custom-code guardrail sandbox escape, May 2026
github.com

Anthropic on the market: 83, 111.50, 97.17

Last issue it was a range of 70 to 78 dollars. In the end the offering was upsized, the price raised, and the stock closed day one 17% above the offer price.

On 18 November Anthropic priced its offering at 83 dollars per Class A share, above the announced range. The offering was upsized from 825 to 980 million shares: 956.35 million sold by the company and 23.65 million by existing shareholders. Underwriters have an option on another 147 million shares. The SEC declared the registration effective the same day.

Trading on Nasdaq under ANT began on 19 November. The stock opened at 111.50 dollars, reached 115.10 during the day, fell to 92.72 and closed at 97.17 dollars. Volume was 857 million shares.

At the opening the co-founders Dario and Daniela Amodei spoke. Daniela Amodei said more than half of Fortune 500 companies use Claude.

DAY ONE, ANT

Offer price
$83.00

Open
$111.50

High
$115.10

Low
$92.72

Close
$97.17

SOURCES

1 **Anthropic**, Anthropic prices initial public offering, 18 Nov 2026 anthropic.com

2 **Nasdaq**, Anthropic Marks its IPO at Nasdaq, 19 Nov 2026 nasdaq.com

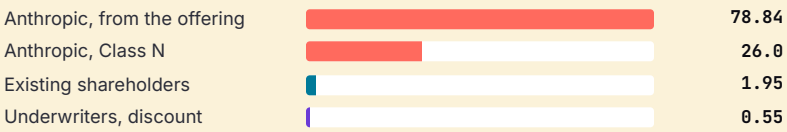
3 **Nasdaq**, ANT historical data, 19 Nov 2026 nasdaq.com

MONEY / 424B4 PROSPECTUS / WHO GOT THE MONEY

Who got the money, and what is only an estimate

Headlines mix four different numbers: the total offering, money for the company, money for existing shareholders and market value. Here are all four, separately.

MONEY FROM THE OFFERING AND PRIVATE PLACEMENT, $BN



Before offering expenses. Per the prospectus, net proceeds to the company are about $104.8bn. Source: SEC EDGAR.

According to the final prospectus of 19 November, the whole offering is worth 81.34 billion dollars. Underwriters received a 550 million discount, Anthropic 78.84 billion before expenses, and selling shareholders 1.95 billion. On top of that, Google, NVIDIA, AMD, Broadcom, Micron, Microsoft and TSMC are buying 26 billion dollars of Class N shares at the offer price. The company estimates net proceeds at about 104.8 billion, or 117 billion if underwriters exercise their option.

After the offering there are about 13.22 billion shares of all classes. Multiplied by the offer price that is about 1.10 trillion dollars, and by the closing price about 1.28 trillion. That is our calculation on the basic share count, excluding 1.77 billion options and 1.03 billion unvested RSUs, so a fully diluted value would be higher. None of these numbers is revenue: that was 33.2 billion for the first nine months of 2026.

NUMBER

104.8 $bn

estimated net proceeds to the company. SEC EDGAR

NUMBER

~1.28 $tn

value at close, our calculation, excl. options. SEC EDGAR

SOURCES

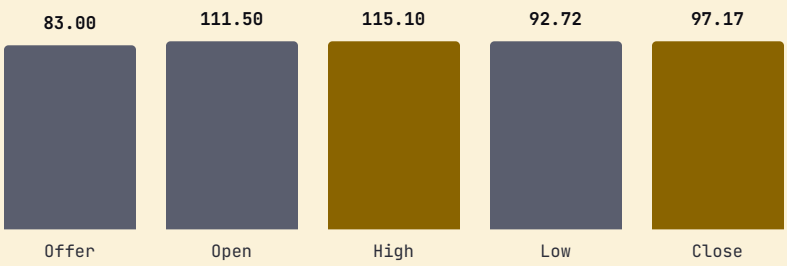
1 **SEC EDGAR**, Anthropic, PBC, Prospectus 424(b)(4), 19 Nov 2026 sec.gov

2 **Nasdaq**, ANT historical data, 19 Nov 2026 nasdaq.com

MONEY / DAY ONE / NASDAQ

From 111.50 to 97.17 dollars in a day

The stock opened 34% above the offer price and closed 13% below the open. Both say something about how the market sees AI.



According to Nasdaq historical data. Source: Nasdaq.

Investors who got shares at the offer price were about 17% up at the end of the day. Those who bought at the open were about 13% down. The gap shows how far demand exceeded supply at the very start, and also that the price quickly moved back toward the level at which the banks allocated the offering.

Last issue we wrote about 900.9 billion dollars in compute commitments and the company's own warning that revenue would not keep growing at the same pace. Those numbers did not change with the listing, but Anthropic now reports publicly, quarterly, under SEC rules.

For our readers: this describes the first trading day, it is not investment advice.

NUMBER

857 M

shares traded on day one. Nasdaq

SOURCES

1 **Nasdaq**, ANT historical data, 19 Nov 2026 nasdaq.com

2 **Nasdaq**, Anthropic Marks its IPO at Nasdaq, 19 Nov 2026 nasdaq.com

3 **Anthropic**, Anthropic prices initial public offering, 18 Nov 2026 anthropic.com



NVIDIA: 115.8 billion and new leadership

The first quarterly report after Jensen Huang's resignation is signed by the finance chief who is temporarily running the company. The numbers are records again.

NVIDIA reported revenue of 115.8 billion dollars for the third quarter of fiscal 2027, which ended on 25 October, up 20% from the previous quarter and 103% from a year earlier. Data Center brought in 107.6 billion. Gross margin was 74.8%, and net income 65.3 billion dollars.

For the fourth quarter, which has an extra week, the company expects 130 billion plus or minus 3% and assumes no data centre compute revenue from China. Gross margin is expected to fall to 71 to 72%.

The Vera Rubin NVL72 platform has started shipping, and NVIDIA calls it the fastest product ramp in its history. Racks are already running at CoreWeave, Google Cloud, Microsoft Azure, Oracle Cloud and Nebius. NVIDIA also says it expanded Anthropic's access to Grace Blackwell and Vera Rubin capacity through cloud providers.

NUMBER

107.6 $bn

Data Center revenue, +110% year on year. NVIDIA

NUMBER

130 $bn

Q4 outlook, excluding China. NVIDIA

SOURCES

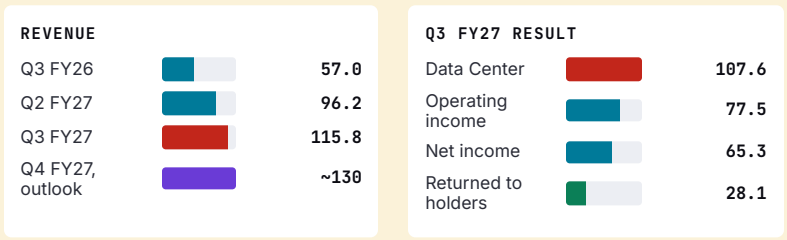
1 **NVIDIA**, NVIDIA Announces Financial Results for Third Quarter Fiscal 2027, 17 Nov 2026 nvidianews.nvidia.com

MONEY / NVIDIA / LEADERSHIP AND RISKS

Record numbers, open questions

NVIDIA is still the best indicator of how much the world spends on AI. This week it also shows how dependent that spending is on politics and people.

NVIDIA BY QUARTER, $BN



GAAP figures from NVIDIA's release. Fiscal 2027 starts in late January 2026. Source: NVIDIA.

The context of the week is a change at the top. On 12 November NVIDIA announced that Jensen Huang had resigned as president, CEO and board member, effective immediately. Colette Kress, CFO since 2013, is running the company on an interim basis and keeps the finance role. Harvey C. Jones became the first board chair in the company's history. The announcement gives no reason for the resignation, and we do not repeat speculation.

On 17 November the board created a standing committee for national security and export controls, with independent experts. Together with an outlook that excludes China, this shows geopolitics has become part of the business plan, not just a footnote risk.

For judging the sustainability of the AI industry, this matters: customers still pay record amounts, but next quarter's margin is falling, and growth depends on a few large cloud providers and AI labs.

DATE

The resignation was announced on 12 November, a day before this issue's period. We cite it as context.

SOURCES

1 **NVIDIA**, NVIDIA Announces Financial Results for Third Quarter Fiscal 2027, 17 Nov 2026 nvidianews.nvidia.com

2 **NVIDIA**, NVIDIA Announces CEO Transition, 12 Nov 2026 nvidianews.nvidia.com

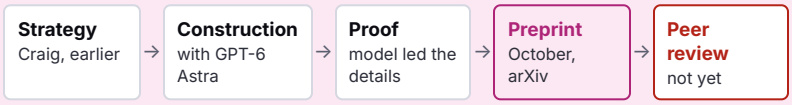
3 **NVIDIA**, NVIDIA Board of Directors Establishes National Security and Export Control Oversight Committee, 17 Nov 2026 nvidianews.nvidia.com

LAB / OPENAI, UCSB / 15 NOV 2026

The Standard Model on a lattice, with AI's help

Physicist Nathaniel Craig describes how he worked with a model on a candidate rigorous definition of the Standard Model. It is a preprint, not a peer-reviewed paper.

STATUS OF THE RESULT



Based on the post on OpenAI's science blog.

The Standard Model describes the known particles and three of the four fundamental forces and agrees superbly with experiment. Yet it has no nonperturbative definition, one that does not treat interactions as small corrections. The usual route is a lattice, a grid of points instead of continuous spacetime, but the Standard Model is a chiral theory: it looks different in a mirror, and that has long been hard to write down on a lattice.

In the first guest post on OpenAI's new science blog, Craig of the University of California, Santa Barbara writes that he had a strategy but could not prove it worked in all the cases needed. In his words, GPT-6 Astra helped turn the strategy into a construction and took the lead in developing the detailed proof, published as a preprint in October.

Craig also lists what remains open, including defining the theory in continuous spacetime. The result still has to pass review by the community.

STATUS

Preprint on arXiv. The post appeared on the blog of the company whose model was used.

SOURCES

1 **OpenAI**, A step toward defining the Standard Model on a lattice, 15 Nov 2026
openai.com

2 **arXiv**, Craig, candidate lattice definition of the Standard Model (2610.09093), Oct 2026
arxiv.org

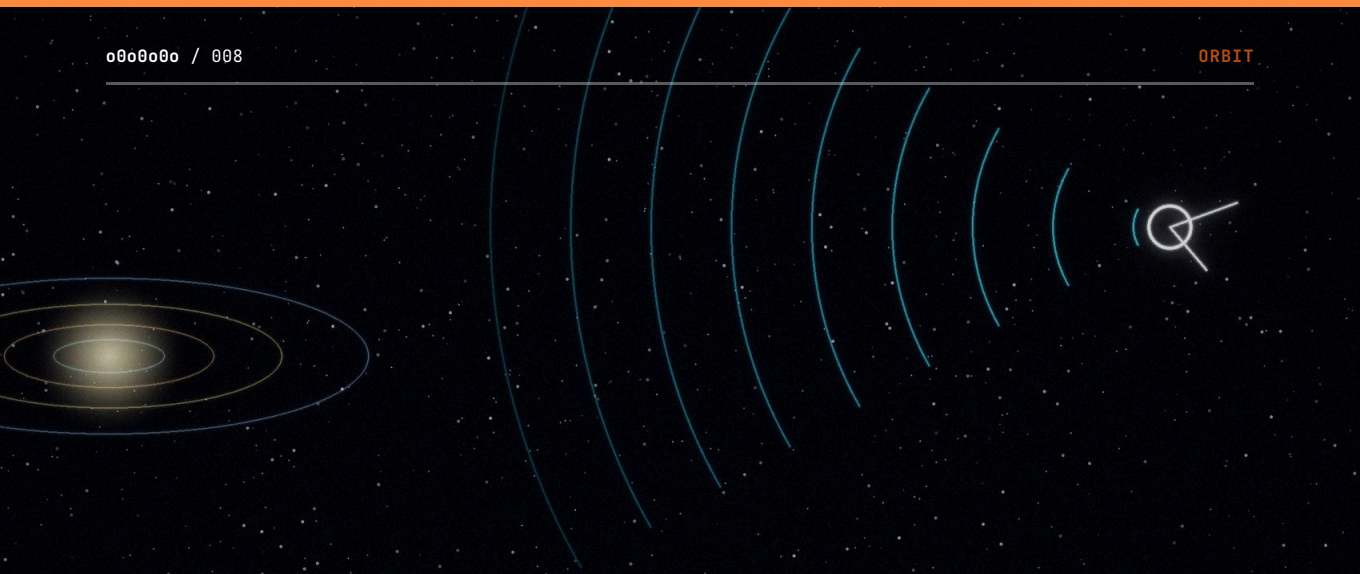


Illustration: a probe far beyond the planets and a signal that travels for 24 hours. 0000000, NOT A PHOTOGRAPH

ORBIT / NASA / 18 AND 19 NOV 2026

Voyager 1: one light-day from home

Last issue it was a preview. On 18 November Voyager 1 reached the distance light travels in 24 hours, after more than 49 years of flight.

According to NASA, Voyager 1 reached one light-day from Earth on 18 November at 2:16 Pacific time, 11:16 Central European time. That is about 25.9 billion kilometres. The JPL team heard from the probe 24 hours later, the time the signal needs to reach Earth.

Launched in 1977, Voyager 1 in 2012 became the first spacecraft to leave the heliosphere, the bubble of particles and magnetic fields created by the Sun, and enter interstellar space. It travels at about 61,000 kilometres per hour.

An important detail in the release: since 14 November the team has received no science or engineering data. The probe is sending its carrier signal, which normally carries that data, and the milestone was confirmed using that signal and trajectory calculations. Engineers are analysing the problem.

VOYAGER 1

Launched
1977.

Distance
~25.9bn km

Signal
24 h one way

Data
none since 14 Nov

SOURCES

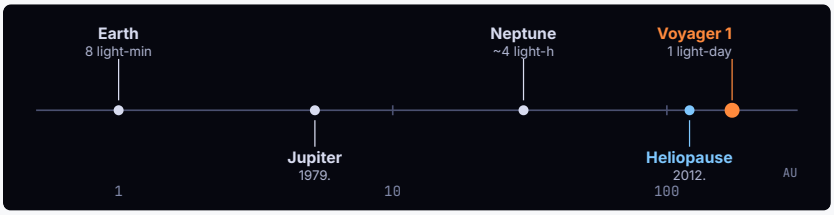
1 **NASA**, NASA's Voyager 1 Makes History Reaching One Light-Day From Earth, 19 Nov 2026
nasa.gov

ORBIT / DEEP SPACE NETWORK / SIGNAL AND POWER

A carrier signal is not the same as data

Antennas on Earth hear Voyager, but are not currently getting what it carries. The difference matters for understanding the mission.

DISTANCE FROM THE SUN, ASTRONOMICAL UNITS, LOG SCALE



One light-day is about 173 AU. The heliopause position is approximate, based on Voyager 1's crossing.

A carrier signal is the radio wave data rides on, like a truck that carries the cargo. When Deep Space Network antennas receive it, the team knows the probe is alive and where it is, but without data it does not know what the measurements show or what state the systems are in.

Communication with Voyager 1 now takes 24 hours one way. The team, says project manager Suzanne Dodd, listens to the probe every day and sends commands once a week. A reply to a command arrives two days later at the earliest.

Power comes from a radioisotope generator almost half a century old, which loses about 4 watts a year. This year engineers replaced three devices with lower-power ones, and NASA expects the instruments to keep running at least until the 50th launch anniversary, next summer.

NUMBER

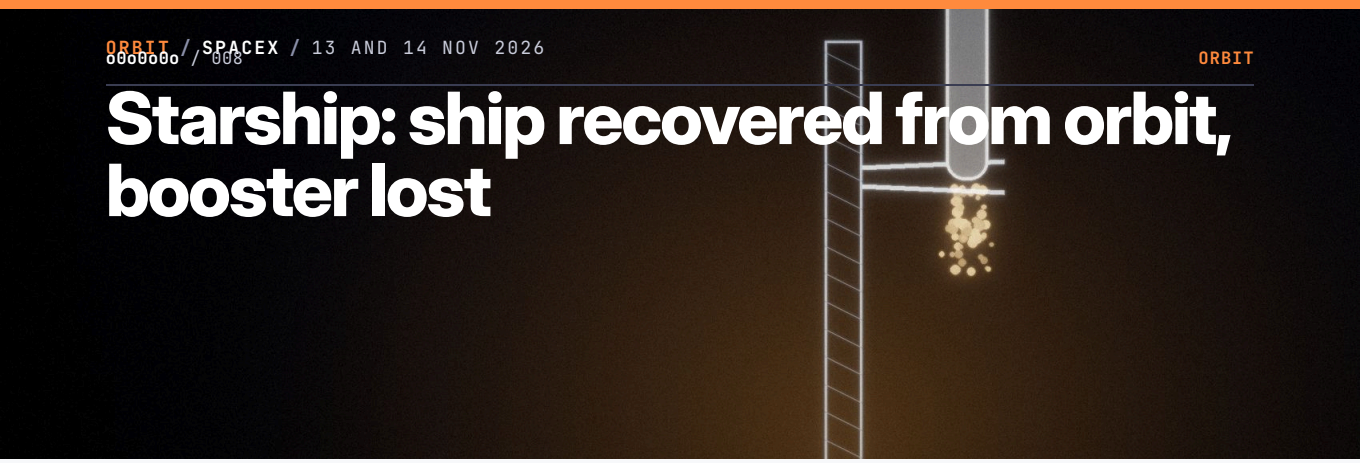
4 W

power lost each year.
NASA

SOURCES

1 **NASA**, NASA's Voyager 1 Makes History Reaching One Light-Day From Earth, 19 Nov 2026
nasa.gov

Starship: ship recovered from orbit, booster lost



SpaceX tried to recover a whole orbital rocket. The upper stage came back after about 22 hours; the booster went into the sea.

Flight 15 lifted off from Starbase in Texas on 13 November at 22:47 UTC. According to Space.com, the ship reached orbit at about 275 kilometres, the second time in the programme's history, and deployed 26 Starlink V3 satellites.

The booster was not caught by the launch tower's arms. It veered offshore and triggered its termination system. SpaceX first cited a guidance issue; a day later Kate Tice said the booster ended its return burn early after a low fuel alarm. The cause is still under investigation.

According to Space.com, the ship came back on 14 November, the first recovery of a Starship upper stage from orbit. Recovering both stages remains the goal of future flights.

FLIGHT 15

Launch
13 Nov 22:47 UTC

Orbit
~275 km

Payload
26 Starlink V3

Booster
lost

Ship
recovered 14 Nov

SOURCES

1 **Space.com**, SpaceX launches Starship on ambitious 22-hour Flight 15 test, 13 Nov 2026
space.com

2 **Space.com**, Why did Starship Flight 15 booster abort its chopsticks catch?, 14 Nov 2026
space.com

3 **Space.com**, Starship Flight 15 live updates: Starship lands from orbit for 1st time, 14 Nov 2026
space.com

ORBIT / ESA EUCLID / 12 NOV 2026

Euclid: 900 million galaxies and 15,000 lens candidates

Last issue we listed DR1-Foundation as a preview with no date. It was released on 12 November.

ESA and the Euclid Consortium released the first part of the first big data release, DR1-Foundation. According to Physics World, it covers about 13 months of observations, about 1,900 square degrees of sky and more than 900 million galaxies.

Alongside it came a catalogue of strong gravitational lenses, led by a team at the University of Portsmouth. It holds about 15,000 high-confidence candidates; previously only a few hundred such systems were known in space-based imaging. These are candidates selected by shape, and none has yet been confirmed by spectroscopy. The catalogue paper has been submitted to Astronomy and Astrophysics.

The full DR1 release, with products for galaxy clustering and weak lensing, is expected in mid-2027.

NUMBER

15,000

strong gravitational lens candidates. University of Portsmouth

SOURCES

1 **Physics World**, ESA's Euclid mission unveils treasure trove of cosmic data, 12 Nov 2026 physicsworld.com

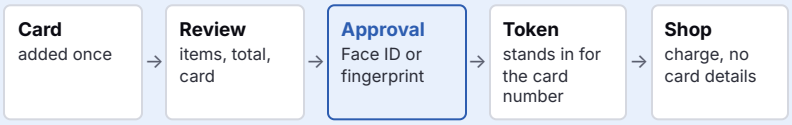
2 **University of Portsmouth**, Portsmouth astronomers help build the biggest ever gravitational lens catalogue, 12 Nov 2026 port.ac.uk

DIGITAL LIFE / OPENAI / 13 NOV 2026

The agent pays, the card stays hidden

ChatGPT Wallet lets an agent buy on your behalf, but you confirm every purchase with biometrics, and neither the agent nor the shop sees the card number.

HOW ONE PURCHASE WORKS



Based on OpenAI's 13 November description.

ChatGPT Wallet was built with American Express, Chase, Mastercard and Visa. You add a card once, and before every purchase you review the items, total and card and approve with a security check, such as Face ID or a fingerprint. The network then issues a token, a substitute for the real card number, so neither the agent nor the shop gets the actual details.

Rewards and dispute protection stay the same as with a normal card purchase. To start, it is available only to ChatGPT Pro subscribers in the US. OpenAI says it will later be possible to pre-approve purchases with limits, categories and time windows.

For users this is a good model: the agent does the work, but money does not move without your explicit approval of every item.

AT A GLANCE

Availability
ChatGPT Pro, US

Partners
Amex, Chase, Mastercard, Visa

Approval
for every purchase

SOURCES

1 **OpenAI**, Introducing ChatGPT Wallet, 13 Nov 2026
openai.com

o0o0o0o / 008

USE THIS

An agent with 200 tools, without 200 schemas per request

This guide combines three of the week's topics: Tool Retrieval from the Gemini API, per-user permissions and protection against a LLMJacking bill.

EXAMPLE CALCULATION FOR ONE REQUEST, WITH ASSUMPTIONS

APPROACH	INPUT TOKENS FOR TOOLS	COST WITH NEON, $
All schemas in the request	200 × 250 = 50,000	0.150
Names, descriptions and 3 loaded schemas	200 × 30 + 3 × 250 = 6,750	**0.020**

Expected, not measured. Assumptions: 250 tokens per schema, 30 per name and description, $3 per million input tokens, no caching. Source: Google, Google.

The table's calculation is an illustration: the real token count depends on your schemas, and prompt prefix caching can change the picture considerably. Google therefore advises measuring savings on your own traffic. The plan for that is at the end of the guide.

The code follows Google's announcement and documentation examples, but the editors have not run it. Results we mention should be read as expected.

SOURCES

1 **Google**, Introducing Tool Retrieval in the Gemini API, 19 Nov 2026
blog.google

2 **Google**, Introducing Gemini 4 Neon, 18 Nov 2026
blog.google

3 **Google Threat Intelligence Group**, GTIG AI Threat Tracker: Agentic Threat Activity Crosses a Significant Threshold, 16 Nov 2026
cloud.google.com

USE THIS / STEP 1 / DEFERRED LOADING

Turn on server-side Tool Retrieval

The simplest variant: you add one tool and one flag, and Google runs the search. The example is in Python, as in the official announcement.

```
from google import genai

client = genai.Client()

tools = [
    {"type": "tool_search"},
    # keep frequent tools loaded, without defer_loading
    {"type": "function", "name": "issue_refund",
     "description": "Refund an order.",
     "defer_loading": True,
     "parameters": {"type": "object", "properties": {
         "order_id": {"type": "string"}, "amount": {"type": "number"}},
         "required": ["order_id", "amount"]}},
    # ... other deferred tools
]

interaction = client.interactions.create(
    model="gemini-4-neon",
    input="Order A-1042 arrived damaged. Please refund 40 euros.",
    tools=tools,
)
for step in interaction.steps:
    print(step.type)  # tool_search_call, tool_search_result, function_call
```

WHAT HAPPENS

The model sees the name and description of issue_refund, asks for it by name, the API loads the schema and the model emits a function_call, all in one call. From there it is business as usual: your code performs the action and returns a function_result.

MCP

The same flag works on an mcp_server tool: every tool from that server is then loaded on demand.

SOURCES

1 **Google**, Introducing Tool Retrieval in the Gemini API, 19 Nov 2026
blog.google

USE THIS / STEP 2 / YOUR OWN REGISTRY

Tool search on your side, with per-user permissions

If tools depend on the user, you run the search yourself, at one extra round trip per search.

```
type ToolDef = {
  name: string; description: string; scopes: string[];
  parameters: Record<string, unknown>;
};

export function searchTools(registry: ToolDef[], query: string,
  userScopes: Set<string>, limit = 5) {
  const words = query.toLowerCase().split(/\s+/).filter(Boolean);
  return registry
    // the user only sees tools they hold every scope for
    .filter((t) => t.scopes.every((s) => userScopes.has(s)))
    .map((t) => {
      const text = `${t.name} ${t.description}`.toLowerCase();
      const score = words.filter((w) => text.includes(w)).length;
      return { t, score };
    })
    .filter((x) => x.score > 0)
    .sort((a, b) => (a.score < b.score ? 1 : -1))
    .slice(0, limit)
    .map(({ t }) => ({ type: "function", name: t.name,
      description: t.description, parameters: t.parameters }));
}
```

HOW IT FITS

The model calls tool_search with "execution": "client", your code runs searchTools and returns the schemas as a function_result.

The key advantage is the permission filter before the model ever sees a schema: a refund tool cannot appear for a user without that permission, whatever the model asks for.

WHY THE FILTER COMES FIRST

A model cannot call a tool whose schema it never received. That is a stronger boundary than an instruction in the prompt.

SOURCES

1 **Google**, Introducing Tool Retrieval in the Gemini API, 19 Nov 2026
blog.google

USE THIS / STEPS 3 AND 4 / BUDGET AND MEASUREMENT

A daily budget and a measurement plan

An agent with many tools and a model key is a target. A service-level budget limits the damage, and measurement shows whether Tool Retrieval really saved anything.

```
const DAILY_LIMIT_USD = 50;
const PRICE = { input: 3 / 1e6, output: 15 / 1e6 }; // Neon, list price
let day = new Date().toISOString().slice(0, 10);
let spent = 0;

export function charge(usage: { input: number; output: number }) {
  const today = new Date().toISOString().slice(0, 10);
  if (today !== day) { day = today; spent = 0; }
  spent += usage.input * PRICE.input + usage.output * PRICE.output;
  if (spent > DAILY_LIMIT_USD) {
    alertOps(`AI budget exceeded: ${spent.toFixed(2)} $`);
    throw new Error("daily AI budget is spent");
  }
}
```

HOW TO MEASURE THE SAVINGS

Run the same set of at least 50 real requests twice: once with all schemas in the request, once with Tool Retrieval. Record input tokens from the response, latency, whether the right tool was chosen and the cost per successfully finished task. We expect fewer input tokens and similar or better tool choice, but that is an expectation, not a measured result.

LIMITATIONS

The feature is in preview and works only in the Interactions API. The in-memory counter here is an example: in production keep it in a database and add a limit at the provider itself, because a stolen key bypasses your service.

WHAT TO RECORD

1. input tokens
2. latency
3. right tool?
4. cost per task

SOURCES

1. **Google**, Introducing Tool Retrieval in the Gemini API, 19 Nov 2026 blog.google
2. **Google Threat Intelligence Group**, GTIG AI Threat Tracker: Agentic Threat Activity Crosses a Significant Threshold, 16 Nov 2026 cloud.google.com
3. **Google**, Introducing Gemini 4 Neon, 18 Nov 2026 blog.google

SHIPPED / 13 NOV 2026 → 20 NOV 2026

Arrived this week

What is actually available, with platform and limits. Announcements with future dates are not on this page.

OPENAI · 17 NOV

GPT-6.2 Astra, Sol

ChatGPT Work, Codex, API and Copilot.

AVAILABLE OpenAI · GitHub Changelog

GOOGLE · 18 NOV

Gemini 4 Neon

API, Antigravity, Gemini app for Pro and Ultra.

ROLLOUT Google

GOOGLE · 19 NOV

Genie 4 in the API

Interactions API and AI Studio. Ultra in Project Genie.

AVAILABLE Google · Google

GOOGLE · 19 NOV

Tool Retrieval

Interactions API only, Gemini 3.6 and newer.

PUBLIC PREVIEW Google

MICROSOFT · 17 NOV

Copilot SDK 2.0

Rust runtime, six languages.

GA Microsoft

MICROSOFT · 17 NOV

Microsoft Perception

Security agents and playbooks.

GA Microsoft

MICROSOFT · 17 NOV

Defender for local agents

Windows, Defender for Endpoint P2.

GA Microsoft

GITHUB · 17 NOV

Secret scanning on demand

Copilot CLI, app, SDK and code review.

PUBLIC PREVIEW GitHub Changelog

VERCEL · 18 NOV

Vercel Board

All plans, experimental.

AVAILABLE Vercel

VERCEL · 19 NOV

Billing CLI

Upcoming invoice estimates need a paid plan.

AVAILABLE Vercel

5 MINUTES / 13 NOV 2026 → 20 NOV 2026

In brief

Short news. If a topic has its own article, the card points to it.

GITHUB · 16 NOV

End of the Local harness

VS Code removes it in February 2027. Page 23.

GitHub Changelog

GITHUB · 17 NOV

Copilot Evals

Repeatable tests for skills, agents and MCP. Page 28.

GitHub Changelog

GITHUB · 17 NOV

Copilot in Agent 365

Copilot activity visible in Microsoft's agent monitoring.

GitHub Changelog

VERCEL · 18 NOV

Free search

Parallel fast mode until 31 January. Page 29.

Vercel

NVIDIA · 17 NOV

Export committee

A new board committee for national security. Page 38.

NVIDIA

OPENAI · 18 NOV

GPT-Synopsys

A new model for chip design tools. Page 11.

OpenAI

MICROSOFT · 17 NOV

Windows Server 2028

In preview.

Microsoft

MICROSOFT · 17 NOV

Cobalt 200

An Arm processor for agents, 24 regions. Page 26.

Microsoft

THIS WEEK IN NUMBERS / 13 NOV 2026 → 20 NOV 2026

The week in numbers

81.85%

GPT-6.2 Astra on the internal research debugging test

OPENAI · 17 NOV

46%

of 2 to 4 h tasks OpenAI agents finish alone

OPENAI · 18 NOV

2,400$

median researcher's daily spend on agents

OPENAI · 18 NOV

13%

less area in the Jalapeño vector unit

OPENAI · 18 NOV

8 / 40$

GPT-6.2 Astra per million tokens, in and out

OPENAI · 17 NOV

3 / 15$

Gemini 4 Neon per million tokens

GOOGLE · 18 NOV

100+

agents in Microsoft's MDASH system

MICROSOFT · 17 NOV

1$M

largest LLMJacking bill before it was stopped

GOOGLE THREAT INTELLIGENCE GROUP · 16 NOV

THIS WEEK IN NUMBERS / 13 NOV 2026 → 20 NOV 2026

More numbers of the week

83$

Anthropic offer price

ANTHROPIC · 18 NOV

97.17$

ANT at the first day's close

NASDAQ · 19 NOV

104.8$bn

estimated net proceeds to Anthropic

SEC EDGAR · 424B4

115.8$bn

NVIDIA quarterly revenue, +103%

NVIDIA · Q3 FY27

7

new CISA KEV entries, one an AI gateway

CISA · 13 TO 18 NOV

25.9bn km

Voyager 1 from Earth, one light-day

NASA · 18 NOV

900M

galaxies in Euclid DR1-Foundation

PHYSICS WORLD · 12 NOV

26

Starlink V3 satellites on Flight 15

SPACE.COM · 13 NOV

NEXT / 13 NOV 2026 → 20 NOV 2026

What we are watching

Future events after the 20 November cut-off. None of them had happened when the issue closed.

→ 20 NOV

Anthropic offering closes

Per the announcement, subject to customary conditions.

Anthropic

→ 21 NOV

KEV deadlines

Synology, REDCap, Apache Axis.

CISA

→ 1 DEC

KEV deadline: Puppet

Puppet Enterprise, for US federal agencies.

CISA

→ 2 DEC

KEV deadline: LiteLLM

The patch has been available since May.

CISA · BerriAI / GitHub

→ DECEMBER

Copilot Code GA

Alongside Agent 365 for third-party agents.

Microsoft

→ 23 DEC

NVIDIA dividend

$0.25 per share.

NVIDIA

→ 31 JAN 2027

End of free search

Parallel fast mode on AI Gateway.

Vercel

→ FEB 2027

VS Code Local harness

The old harness is removed.

GitHub Changelog

→ MID-2027

Euclid, full DR1

Products for clustering and weak lensing.

Physics World

→ SUMMER 2027

Voyager at 50

NASA expects the instruments to work at least until then.

NASA

SOURCES / 13 NOV 2026 → 20 NOV 2026

All sources

Sources are grouped by page, in order of first appearance. The primary source comes first. Every link is live.

04 EDITORIAL: WHEN THE TOOL STARTS BUILDING THE TOOL

1. **OpenAI**. GPT-6.2 System Card. 17 Nov 2026 deploymentsafety.openai.com/gpt-6-2-astra-sol
2. **OpenAI**. Research acceleration: The jagged frontier. 18 Nov 2026 openai.com/index/research-acceleration-jagged-frontier/
3. **Nasdaq**. ANT historical data. 19 Nov 2026 nasdaq.com/market-activity/stocks/ant/historical
4. **OpenAI**. Advancing chip design with OpenAI models. 18 Nov 2026 openai.com/index/advancing-chip-design-with-our-models/

06 COVER STORY: THE FIRST PUBLIC MODEL AT THE SELF-IMPROVEMENT THRESHOLD

5. **OpenAI**. Introducing GPT-6.2 Astra and Sol. 17 Nov 2026 openai.com/index/introducing-gpt-6-2-astra-sol/
6. **OpenAI**. Control Room. 9 Nov 2026 openai.com/index/control-room/

13 MODEL WATCH: GPT-6.2 ASTRA AND SOL: FEWER TOKENS FOR THE SAME WORK

7. **GitHub Changelog**. OpenAI GPT-6.2 Astra and GPT-6.2 Sol now available. 17 Nov 2026 github.blog/changelog/2026-11-17-openais-gpt-6-2-astra-and-gpt-…

15 MODEL WATCH: GEMINI 4 NEON: CARBON QUALITY, FASTER AND CHEAPER

8. **Google**. Introducing Gemini 4 Neon. 18 Nov 2026 blog.google/innovation-and-ai/models-and-research/gemini-models…
9. **Google DeepMind**. Gemini 4 Neon, Model Card. 18 Nov 2026 deepmind.google/models/model-cards/gemini-4-neon/
10. **GitHub Changelog**. Gemini 4 Neon in GitHub Copilot. 18 Nov 2026 github.blog/changelog/2026-11-18-gemini-4-neon-in-github-copilot
11. **Vercel**. Gemini 4 Neon now available on AI Gateway. 18 Nov 2026 vercel.com/changelog/gemini-4-neon-now-available-on-ai-gateway

16 MODEL WATCH: ASTRA, NEON, FABLE: A COMPARISON WITHOUT A WINNER

12. **GitHub Changelog**. GitHub Copilot Fusion updates: more availability, benchmarks, and models. 17 Nov 2026 github.blog/changelog/2026-11-17-github-copilot-fusion-updates-…

17 MODEL WATCH: MODELS OF THE WEEK

13. **Google**. Genie 4 is now available. 18 Nov 2026 blog.google/innovation-and-ai/models-and-research/google-deepmi…
14. **Google**. Genie 4 is now available on the Gemini API. 19 Nov 2026 blog.google/innovation-and-ai/technology/developers-tools/genie…
15. **Microsoft**. Microsoft Ignite 2026 Book of News. 17 Nov 2026 news.microsoft.com/events/ignite/2026/book-of-news
16. **GitHub Changelog**. MAI-Code-2 and MAI-Code-2-Flash in GitHub Copilot. 17 Nov 2026 github.blog/changelog/2026-11-17-mai-code-2-and-mai-code-2-flas…

18 MODEL WATCH: GENIE 4: A WORLD THAT CHANGES AS YOU WALK THROUGH IT

17. **Google DeepMind**. Genie 4: A new generation of world simulation. 27 Oct 2026 deepmind.google/blog/genie-4-a-new-generation-of-world-simulati…

SOURCES / CONTINUED

21 THE BIG SHIFT: COPILOT BECOMES A WORK PLATFORM

18. **Microsoft**. Microsoft Ignite 2026: Empowering human ambition, intelligence and trust. 17 Nov 2026 news.microsoft.com/source/2026/11/17/microsoft-ignite-2026-empo…

23 THE BIG SHIFT: WHAT DEVELOPERS CAN USE RIGHT AWAY

19. **GitHub Changelog**. VS Code retires the Local harness in February 2027. 16 Nov 2026 github.blog/changelog/2026-11-16-vs-code-retires-the-local-harn…
20. **GitHub Changelog**. Enterprise Live Migrations from Azure DevOps generally available. 19 Nov 2026 github.blog/changelog/2026-11-19-enterprise-live-migrations-fro…

27 BUILD: TOOL RETRIEVAL: AN AGENT WITH A HUNDRED TOOLS

21. **Google**. Introducing Tool Retrieval in the Gemini API. 19 Nov 2026 blog.google/innovation-and-ai/technology/developers-tools/gemin…

28 BUILD: GITHUB: FUSION ON A PUBLIC LEADERBOARD, SECRETS BEFORE MERGE

22. **GitHub Changelog**. Secret scanning on demand for developers and agents. 17 Nov 2026 github.blog/changelog/2026-11-17-secret-scanning-on-demand-for-…
23. **GitHub Changelog**. GitHub Copilot Evals is now in public preview. 17 Nov 2026 github.blog/changelog/2026-11-17-github-copilot-evals-is-now-in…
24. **GitHub Changelog**. GitHub Copilot models deprecated on November 16, 2026. 16 Nov 2026 github.blog/changelog/2026-11-16-github-copilot-models-deprecat…
25. **GitHub Changelog**. Red Hat Enterprise Linux runner images now generally available. 19 Nov 2026 github.blog/changelog/2026-11-19-red-hat-enterprise-linux-runne…

29 BUILD: VERCEL: A MESSAGE BOARD FOR AGENTS AND BILLING IN THE TERMINAL

26. **Vercel**. Vercel Drives now reads files up to 6x faster. 18 Nov 2026 vercel.com/changelog/vercel-drives-now-reads-files-up-to-6x-fas…
27. **Vercel**. Vercel Board is now available. 18 Nov 2026 vercel.com/changelog/vercel-board-is-now-available
28. **Vercel**. Vercel Billing CLI is now available. 19 Nov 2026 vercel.com/changelog/vercel-billing-cli-is-now-available
29. **Vercel**. Parallel web search is free on AI Gateway and eve through January 31. 18 Nov 2026 vercel.com/changelog/parallel-web-search-is-free-on-ai-gateway-…

30 SECURITY: ATTACKS AT MACHINE SPEED

30. **Google Threat Intelligence Group**. GTIG AI Threat Tracker: Agentic Threat Activity Crosses a Significant Threshold. 16 Nov 2026 cloud.google.com/blog/topics/threat-intelligence/agentic-threat…

SOURCES / CONTINUED

32 PATCH NOW: PATCH IMMEDIATELY

31. **CISA**. Known Exploited Vulnerabilities Catalog. 18 Nov 2026 cisa.gov/known-exploited-vulnerabilities-catalog
32. **BerriAI / GitHub**. LiteLLM v1.83.10-stable, fix for the custom-code guardrail sandbox escape. May 2026 github.com/BerriAI/litellm/releases/tag/v1.83.10-stable
33. **CISA**. Known Exploited Vulnerabilities, JSON feed, version 2026.11.18. 18 Nov 2026 cisa.gov/sites/default/files/feeds/known_exploited_vulnerabilit…
34. **BleepingComputer**. Check Point warns of zero-day flaw exploited for root access. 13 Nov 2026 bleepingcomputer.com/news/security/check-point-patches-zero-day…
35. **Perforce**. Puppet Enterprise report deserialization code execution. 17 Nov 2026 portal.perforce.com/s/cve/a91Qi000003IaHpIAK/report-deserializa…

34 MONEY: ANTHROPIC ON THE MARKET: 83, 111.50, 97.17

36. **Anthropic**. Anthropic prices initial public offering. 18 Nov 2026 anthropic.com/news/anthropic-prices-initial-public-offering
37. **Nasdaq**. Anthropic Marks its IPO at Nasdaq. 19 Nov 2026 nasdaq.com/newsroom/anthropic-marks-ipo-at-nasdaq

35 MONEY: WHO GOT THE MONEY, AND WHAT IS ONLY AN ESTIMATE

38. **SEC EDGAR**. Anthropic, PBC, Prospectus 424(b)(4). 19 Nov 2026 sec.gov/Archives/edgar/data/2130357/000162828026074944/anthropi…

37 MONEY: NVIDIA: 115.8 BILLION AND NEW LEADERSHIP

39. **NVIDIA**. NVIDIA Announces Financial Results for Third Quarter Fiscal 2027. 17 Nov 2026 nvidianews.nvidia.com/news/nvidia-announces-financial-results-f…

38 MONEY: RECORD NUMBERS, OPEN QUESTIONS

40. **NVIDIA**. NVIDIA Announces CEO Transition. 12 Nov 2026 nvidianews.nvidia.com/news/nvidia-announces-ceo-transition
41. **NVIDIA**. NVIDIA Board of Directors Establishes National Security and Export Control Oversight Committee. 17 Nov 2026 nvidianews.nvidia.com/news/nvidia-board-of-directors-establishe…

39 LAB: THE STANDARD MODEL ON A LATTICE, WITH AI'S HELP

42. **OpenAI**. A step toward defining the Standard Model on a lattice. 15 Nov 2026 openai.com/index/step-toward-defining-the-standard-model/
43. **arXiv**. Craig, candidate lattice definition of the Standard Model (2610.09093). Oct 2026 arxiv.org/abs/2610.09093

40 ORBIT: ONE LIGHT-DAY

44. **NASA**. NASA's Voyager 1 Makes History Reaching One Light-Day From Earth. 19 Nov 2026 nasa.gov/news-release/nasas-voyager-1-makes-history-reaching-on…

42 ORBIT: STARSHIP: SHIP RECOVERED FROM ORBIT, BOOSTER LOST

45. **Space.com**. SpaceX launches Starship on ambitious 22-hour Flight 15 test. 13 Nov 2026 space.com/space-exploration/launches-spacecraft/spacex-starship…
46. **Space.com**. Why did Starship Flight 15 booster abort its chopsticks catch?. 14 Nov 2026 space.com/space-exploration/launches-spacecraft/spacex-says-sta…
47. **Space.com**. Starship Flight 15 live updates: Starship lands from orbit for 1st time. 14 Nov 2026 space.com/news/live/spacex-starship-flight-15-live-updates-nov-…

43 ORBIT: EUCLID: 900 MILLION GALAXIES AND 15,000 LENS CANDIDATES

48. **University of Portsmouth**. Portsmouth astronomers help build the biggest ever gravitational lens catalogue. 12 Nov 2026 port.ac.uk/news-events-and-blogs/news/portsmouth-astronomers-he…
49. **Physics World**. ESA's Euclid mission unveils treasure trove of cosmic data. 12 Nov 2026 physicsworld.com/a/esas-euclid-mission-unveils-treasure-trove-o…

44 DIGITAL LIFE: THE AGENT PAYS, THE CARD STAYS HIDDEN

50. **OpenAI**. Introducing ChatGPT Wallet. 13 Nov 2026 openai.com/index/introducing-chatgpt-wallet/

50 5 MINUTES: IN BRIEF

51. **GitHub Changelog**. Monitor GitHub Copilot activity through Microsoft Agent 365. 17 Nov 2026 github.blog/changelog/2026-11-17-monitor-github-copilot-activit…

o0o0o0o

/ 008

Issue eight. A weekly technology review, free, on tutorijali.app.

Issue 008
Date 20 Nov 2026
Period 13 Nov 2026, 05:37 → 20 Nov 2026, 12:00
Web tutorijali.app

tutorijali.app