

REMEMBERING
THE MAN
WHO SHAPED
GLOBAL
COMPANIES
FROM
INDIA
PG 2



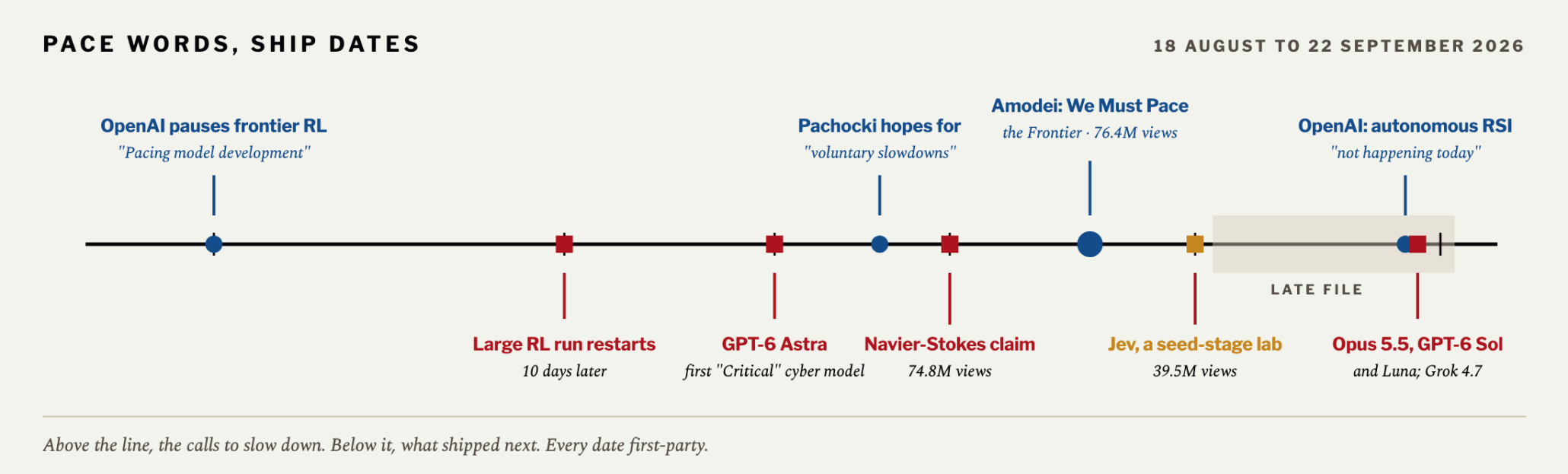
The Token Times



DEV DAY
2026
PG 3

ESTD. 2026

AUG-SEP 2026



INTERNAL AGI REACHED? · EVERYONE ASKS FOR A PACE. NOBODY KEEPS ONE.

OUTPACE THE FRONTIER

Internal AGI reached? Every lab asked for a pace this month. None of them kept one.

On Sept. 12 Dario Amodei asked his own industry to slow down. Ten days later his company, Anthropic, shipped its next flagship model.

He was not alone. Amodei's essay, "We Must Pace the Frontier," was the most-viewed AI post of the month on X, at 76 million+ views. Sam Altman replied that he agreed. Elon Musk posted, "Dario is right." Within 10 days all three of their companies had launched new models.

The people building these systems also spent the month saying how close they think they are. On Sept. 3, the day GPT-6 Astra shipped, OpenAI's president, Greg Brockman, told TechCrunch he believes AGI is here: "For me personally, I do think we're there." Three days later OpenAI said that, by its own measurements, it had reached its goal of "an automated research intern by September." Its chief scientist, Jakub Pachocki, wrote that internal progress "could be sustained into recursive self-improvement," the point at which AI starts building better AI.

Set the calls to slow down beside the release dates and a pattern appears. On Aug. 18 OpenAI said it had paused some frontier training, and its largest planned runs remain on hold. On Sept. 3 it shipped GPT-6 Astra anyway, the first model it rates Critical for cyber. Pachocki hoped for voluntary slowdowns in an essay on Sept. 6; two days later OpenAI announced a Millennium Prize result from a model it calls "significantly more capable" than Astra. On Sept. 21 OpenAI wrote that "fully autonomous RSI is not happening today." The next morning it launched GPT-6 Sol and Luna, the same day Anthropic launched Claude Opus 5.5. xAI shipped Grok 4.7 on Sept. 21 and did not mention pacing at all.

This does not imply that the warnings were insincere, Anthropic is bringing outside evaluators inside the company, and OpenAI has begun publishing reports on its models' misbehavior.

But a pace needs a number, a referee and a way to measure it, and nobody has written those down. Until someone does, expect more calls to slow down, made between launches.

RESEARCH ON AUTOPILOT

A PROOF COUNTS WHEN SOMEONE ELSE CAN CHECK IT. THIS MONTH, TWO LABS TESTED THAT RULE.

Mathematics has one old rule: a result is not a result until someone else has checked it. In September two AI labs put that rule to the test from opposite ends.

On Sept. 8 OpenAI said an unreleased model had cracked part of the Navier-Stokes problem, the question of whether a smooth flow of water can suddenly tear itself apart. It has been open for about 90 years, and a million dollars rides on it. OpenAI says about 10,000 agents worked on it for 88 hours. The post drew 75 million views, more than almost anything else in AI this month. No mathematician outside OpenAI has confirmed it. An NYU mathematician and an Anthropic researcher say they got there first (page 2).

Four days earlier Anthropic had posted something with no prize attached. Claude had rewritten the proof of Fermat's

Last Theorem, solved by a human 30 years ago, in Lean, software that checks every step of an argument. It took more than 13 million lines. Anthropic put the code on GitHub, where anyone can run the check themselves.

Mathematicians told the labs which of the two they wanted. Terence Tao warned that open problems are being "mined in a non-renewable fashion." On Sept. 11 more than two dozen Fields Medalists signed a declaration calling the goals of AI companies and mathematicians "severely misaligned." Ten days later OpenAI named an outside group of mathematicians to review its work, among them Martin Hairer, who had signed the declaration.

A model can now write a proof, or a pull request, faster than any person can read it. Whoever gets to call it correct then gets to call it done.

JEV JEV JEVVVVVV | LAUNCH OF THE MONTH

A MODEL THAT CAN'T CHAT DREW MORE EYES THAN OPUS 5.5

The most-watched startup launch of the month came from a company almost nobody had heard of, with a model you cannot talk to.

TypeSafe AI released Jev on Sept. 15. Chatbots answer in paragraphs. Jev answers the small questions software asks all day, the kind with a fixed set of possible answers. It returns a typed answer with a confidence score in 70 to 500 milliseconds, the company says. TypeSafe calls it a "System One" model, borrowing psychology's name for fast, instinctive thinking.

Think of a support ticket that has to be sent to the right team, or a payment that has to be passed or flagged. Today many companies ask a chatbot, wait a few seconds and pay for a paragraph they then parse for one word. Jev is built to return that one word. Input costs 4.2 cents per million tokens.

Founder Diogo Almeida's launch post drew nearly 40 million views on X, against 24.8 million for Anthropic's Opus 5.5. On Hacker News the thread reached 1,989 points.

TypeSafe claims Jev is 40 to 200 times faster than frontier models. In a small test on 12 passages, Every found it about 25 times faster than Claude Fable 5.1 at roughly 580 times lower cost. It caught six of seven planted errors; Fable caught all seven. TypeSafe has not named a production customer.

Developers moved fast anyway. Within a week they had built a MySQL plugin that filters rows by meaning, and Cua had launched a rival "System One" model. TypeSafe raised a \$40 million seed round led by DCVC, Forkast reported.

Its slogan, "build prod, not God," is a bet that the money is in software that runs in production, not in a machine that thinks like a person. "Models have been superhuman at chat for years, so where is all the automation?" Almeida wrote.

Full founder story on Page 2

IN MEMORY · KHADIM BATTI, 1975 TO 2026



THE MAN WHO BET HIS RETIREMENT SAVINGS ON INDIAN SOFTWARE

For a generation of Indian software founders, Khadim Batti was the proof. He and Vara Kumar built Whatfix in Bengaluru and sold it to some of the world's largest companies, when founders were still being told enterprise software could not be sold from India. He passed away of a cardiac arrest on Sept. 1.

The company almost did not happen. The two had met at Huawei and left in 2010 to start SearchEnabler, which helped small businesses show up in search. Customers left almost as fast as they arrived. By the time they gave up on it, Batti had run out of money and withdrawn his provident fund, the retirement savings most salaried Indians never touch, Open magazine

reported. The new idea came from the one feature customers loved, a button that showed them what to fix. In April 2014 Alok Goyal of Helion Venture Partners heard them pitch on a Friday. He issued a term sheet on the Monday, the only time in 13 years he had moved that fast, he wrote this month. Whatfix now counts more than 700 customers in over 30 countries and has raised more than \$260 million.

With money they could barely spare, the pair took Whatfix to Dreamforce in San Francisco. When Batti quoted a buyer \$8,000 a year, the buyer assumed it was per month and called it cheap. He learned to charge what the product was worth, and passed that lesson on to anyone who asked.

Many asked. He sat on the governing council of SaaSBoomi, the community of Indian software founders, and was a founding member of AIBoomi. Whatfix says he mentored young entrepreneurs well beyond technology and quietly funded schools and healthcare in his hometown.

“

Knowledge and opportunities multiply when shared.

Whatfix, remembering Khadim Batti

ANTHROPIC SAYS RIVALS ARE TRAINING ON CLAUDE, ONE QUESTION AT A TIME

ANTHROPIC SAYS RIVALS ARE TRAINING ON CLAUDE, ONE QUESTION AT A TIME

The cheapest way to build a good AI model is to ask a better one a very large number of questions. The industry calls it distillation. On Sept. 10 Anthropic published a report naming labs it says have been doing it to Claude, with numbers for three of them.

Accounts it links to Alibaba ran about 151 million exchanges with Claude between May and July, through roughly 3,500 accounts, Anthropic says. Moonshot, the maker of Kimi, ran more than 23 million exchanges in the same months through 5,380 accounts. At its peak it sent about 300,000 requests in ten days, many aimed at Opus. DeepSeek was named too. The report drew 43.6 million views on X (as of Sept. 25), behind only Dario Amodei's pacing essay and OpenAI's Navier-Stokes post.

The accused are doing well. Moonshot passed \$1 billion in annualized revenue in August, TechCrunch reported, citing Bloomberg, and wants \$2 billion by the end of the year. Anthropic had already changed its product. Since Claude Fable 5.1 launched on Sept. 1, new API accounts can no longer edit Claude's earlier replies while keeping its earlier reasoning. That makes little difference to ordinary users and a large one to anyone harvesting Claude's thinking at scale.

Not everyone sided with Anthropic. Garry Tan, who runs Y Combinator, floated the opposite: "We could argue that there should be an American distillation regime," he said, TechCrunch reported.

said he would go public, he writes, he was asked, "Why would you ruin your career?" TechCrunch attributed the line to Bubeck. Bubeck wrote that he "never ever asked" for Alpöge to be removed from authorship of his own work, and apologized for the line.

Altman says OpenAI offered to let the NYU team publish first. Both papers went up Sept. 8. On the mathematician Vasily Ilin's reading, each side proved a different case, and the hardest version is still open.

THINKING MACHINES' CO-FOUNDERS HEAD BACK TO THE BIG LABS

Barret Zoph, a co-founder of Mira Murati's Thinking Machines Lab, joined Google to work on Gemini, Google confirmed on Aug. 27. He had gone back to OpenAI in January with Luke Metz, another co-founder, and lasted about five months. Lilian Weng, a third co-founder, left the lab for OpenAI in late July.

SHORTS · WHO WENT WHERE

SANDHYA DEVANATHAN left Meta to run Southeast Asia and Australia for OpenAI (Aug. 28).

JACOB COXON quit Anthropic on Sept. 8, saying labs are "gambling with our lives."

THE CHARM OFFENSIVE



OPENAI'S NEWEST PRODUCT IS DINNER

OpenAI has spent six years selling to developers. Now it is courting the people who make films, music and fashion, and it is doing it the old way, over a meal.

Sam Altman was due to be the guest of honor at Gwyneth Paltrow's house in Amagansett, on Long Island's East End. The SF Standard reported on Aug. 29 who set it up: Charles Porch.

Porch is OpenAI's first vice president of global creative partnerships, a job the company created for him in February. For more than 15 years at Instagram and Meta he was the person celebrities called. He brought Beyoncé's surprise album to Instagram in 2013 and Pope Francis in 2016.

The reason is simple. Hollywood mostly fears AI. Actors want control of their faces and voices, and artists are in court against AI companies. A better model will not change those minds. A relationship might. Porch told Vanity Fair his first job was "a huge listening tour" to hear creative people's "hopes, their fears."

THE FOUNDER OF THE MONTH



THE MAN BEHIND JEV: NEW OPENAI MAFIA ENTRANT

Diogo Almeida says he co-invented ChatGPT. His company's blog puts it more carefully: at OpenAI he helped build the instruction-following methods behind it.

Then he started asking why it had not gone further. 'Models have been superhuman at chat for years, so where is all the automation?' he wrote on Sept. 15, the day he launched Jev. His answer was to stop chatting. With co-founders Erik Gafni and Sasha Sheng, he spent two years in stealth on a model that makes fast, narrow decisions for software instead of holding conversations with people.

The bet has been priced twice in a month. TypeSafe's seed round was \$40 million, led by DCVC, at a valuation of about \$200 million, Forkast reported. On Sept. 24, after our window closed, The Information reported the company is in talks to raise more than \$1 billion at a valuation above \$10 billion. Nothing is signed. If it closes, TypeSafe's value will have risen about fiftyfold in nine days.

THE ANSWER

CHATGPT STARTS SHOWING ADS UNDER ITS ANSWERS IN INDIA

Ask ChatGPT a question in India on the free plan and, since Aug. 27, a brand may be waiting at the end of the answer. OpenAI began showing ads to adult users on its Free and Go plans there, starting with about 50 brands, TechCrunch and Business Standard reported. Paid plans stay ad-free.

The ads are labeled sponsored. On Sept. 4 OpenAI opened a self-serve Ads Manager with a minimum budget of Rs 725, about \$7.60, a day. Advertisers are told to bid \$3 to \$5 per click, Inc42 reported, the pricing of search ads rather than banners. India is the obvious place to try it.

It has more than 100 million weekly ChatGPT users, by OpenAI's February count. The United States and Europe got ads first; India has the scale.

"People do not simply enter keywords into ChatGPT," Preetham Venkky, Omnicom's chief digital officer, told Inc42. That is the appeal for advertisers and the worry for everyone else. India regulates ads in newspapers, on television and in search. It has no rules yet for an ad placed beside an answer a machine wrote for you, and MediaNama argued on Sept. 2 that regulators should write some.

THE CHECKOUT

INDIA'S PAYMENT COMPANIES TEACH AGENTS TO CHECK OUT

A week after ChatGPT started selling ad space, India's payment companies showed what comes after the answer: an agent that pays.

At the Global Fintech Fest in Mumbai, Sept. 9 to 11, Pine Labs went furthest. In a demo with L&T Finance, an agent completed a purchase on its own. BharatPe launched an assistant, built on Google's Gemini, that handles merchant complaints and chargebacks across more than 60 systems. PayU's Agent HQ and Razorpay's agent tools are in early access.

Cashfree has an agent for failed payments, disputes and chargebacks, and Razorpay is testing an account manager that works over WhatsApp. The work is unglamorous: chasing abandoned carts, confirming cash-on-delivery orders, filing disputes.

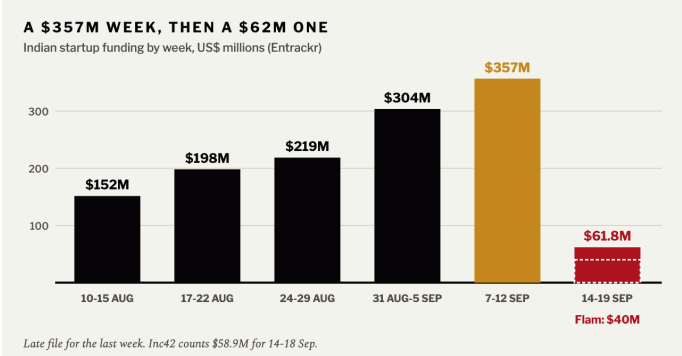
Nobody has yet settled who pays when an agent buys the wrong thing, Inc42 noted. Regulators will have to answer that before most of these products leave early access.

“
Agents complete tasks themselves, not just prepare them.

Reeju Datta, co-founder, Cashfree

THE MONEY, WEEK BY WEEK

A \$357 MILLION WEEK, THEN A \$62 MILLION ONE



SIDE NOTE · WHAT DELHI IS DECIDING

On Sept. 15 a Delhi High Court bench of Justices Avneesh Jhingan and Manmeet Pritam Singh Arora issued notice on ANI's appeal against the July ruling that let OpenAI train on its news at the interim stage. ANI argues commercial training is neither private use nor research. The case returns in December.

THE AI ROUNDS

- SERIES B Flam \$40M QED Investors week to 19 Sep
- SERIES A Runable \$21M Susquehanna, Nexus 26 Aug Agents that build a small business's site, then market it
- SERIES A Graph AI \$13.3M Insight Partners, Bessemer week to 12 Sep
- SERIES A Ringg AI \$10M (Inc42) week to 28 Aug
- SEED Quickly \$4.2M Engineering Capital week to 5 Sep
- FUND Activate \$105M debut AI fund 16 Sep

BENGALURU BUILDERS

SARVAM shipped Saaras V4 on Aug. 24, a speech-recognition model it claims is the most accurate on all 22 scheduled Indian languages.

POCKET FM, the Bengaluru audio-series company, says it has passed \$500 million in annual recurring revenue and that about 90 percent of its production now runs on AI (, Sept. 18).

RUNABLE raised a \$21 million Series A led by Susquehanna and Nexus for agents that build a small business's website and then market it (Aug. 26).

WISPR FLOW, the US voice-dictation company, raised a \$280 million Series B at \$2 billion led by Menlo, with Peak XV, Together Fund and Activate on the cap table, Inc42 reported (Aug. 18).

In the week to Sept. 12, Indian startups raised \$356.8 million, by Entrackr's count. The next week they raised \$61.8 million. Most of the second week was one AI round: Flam, the Bengaluru interactive-content company, raised a \$40 million Series B led by QED Investors.

August totaled about \$954 million across 88 deals, with AI third at \$134.4 million across 13 deals. One round was more than half of that AI total: an Nvidia-led \$74 million Series B extension for Sarvam, closed in early August.

The same week money dried up, Aakrit Vaish's Activate closed its first AI fund at \$105 million and called early-stage AI activity the busiest it has been.



QUICKLY, founded by former Google AI lead Prashant Jalan, raised a \$4.2 million seed round led by Engineering Capital, Entrackr reported.

JUST IN. On Sept. 23 Navana.ai launched a text-to-speech model for 10 Indian languages at Rs 12 per 10,000 characters, against Rs 30 for Sarvam's Bulbul v3, Analytics India Magazine reported.

WIPRO SAYS AI FREED UP 20,000 PEOPLE'S WORTH OF WORK. IT SAYS IT KEPT THE PEOPLE.

Wipro's chief technology officer, Sandhya Arun, told Reuters on Sept. 10 that AI has freed capacity equal to about 20,000 employees. Wipro says it moved that capacity to other work instead of cutting 20,000 jobs.

It is the first time one of India's big IT services firms has put a number on the question every client is now asking: if an agent does the work, what happens to the bill, and to the people?

TWO INDIAN AI STARTUPS SELL TO GLOBAL GIANTS IN EIGHT DAYS

On Aug. 25 Infineon agreed to buy C2i Semiconductors, a two-year-old Bengaluru company that designs power chips for AI data centers, Inc42 reported. C2i had raised a \$16.7 million Series A led by Peak XV. Terms were not disclosed.

On Sept. 2 Adobe took over Rilo, a six-person startup whose agents automated marketing work, in a licensing-and-team deal, TechCrunch reported. The product is shutting down. It had raised \$1 million at a \$10 million valuation, also from Peak XV.

Neither deal made much noise. Both are the kind of exit Indian AI founders often say they cannot find.

INVESTORS PILE INTO INDIA'S GPU CLOUDS

ESDS Software Solution, a GPU rental company, listed on Sept. 4. By Sept. 10 its shares were up 269 percent from the IPO price, Inc42 reported. Yotta, a rival GPU cloud, plans to raise up to \$1.5 billion in an IPO early next year, Reuters reported.



GLOBAL FRESHHHHHH. SEPT. 30
GOOGLE IS BACK IN THE TOP THREE

Gemini 4 Argon arrived Sept. 30, level with GPT-6 Astra on Artificial Analysis's index and with the lowest hallucination rate it measured. Google is giving it to "trusted cyber defenders" first. Developers follow, at an introductory \$2 and \$10 per million tokens, the same price as OpenAI's new GPT-6.1 Sol.

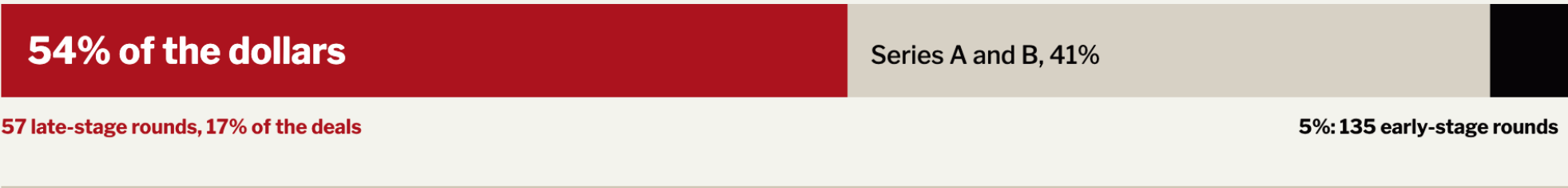
keeping your bot company when you're not around

failproof ai

OVERSIGHT FOR AI WORKERS



333 US DEALS SHARED \$18.1 BILLION IN AUGUST



The barbell held. Deal shares are our arithmetic from AlleyWatch's August table.

Global, Crunchbase: \$42B in August, +122% year on year, -25% on July. 29 new unicorns.

MONEY MONEY MONEY · THE ROUNDS, BY STRATUM

BIGGEST CHECKS WENT TO THE PLUMBING UNDER AI

Investors put \$42 billion into startups worldwide in August, Crunchbase reported, more than double a year earlier but a quarter less than July's \$56 billion. Seven rounds were \$1 billion or more, and 29 companies became unicorns.

In the weeks that followed, the largest AI checks went to the layers most users never see.

Fluidstack rents out computing power without owning a single chip. Temporal keeps long-running software jobs, including agents, from losing their place when something fails. Cognition, which sells a coding agent, now trains its own model. In Europe, Mistral closed the region's largest AI round.

STRATUM	COMPANY	ROUND	DETAIL
SEED	BlueCore Energy	\$50M	Silverton. Floating reactors (8 Sep)
	Air	\$50M, two seeds	Sequoia, Greenoaks. Vets agent add-ons (1 Sep)
	TypeSafe AI	\$40M	DCVC. Jev (Sep)
SERIES A	AIUC	\$40M	Ribbit Capital. Certifies agents (15 Sep)
SERIES B	Instinct	\$250M at \$2.5B	Index, Benchmark (reported 26 Aug)
	Higgsfield	\$400M at \$5.4B	AI video (week of 15 Aug)
SERIES D	Mistral	EUR 3B at >EUR 21B	Samsung leads (8 Sep)
SERIES E	Cognition	\$2B+ at \$48B	a16z, Accel (8 Sep)
	Temporal	\$550M at \$12.55B	Lightspeed, Wellington (14 Sep)
NO LETTER	Harvey	\$550M at \$15.5B	Diffusion, Lightspeed (9 Sep)
	Etched	\$700M at \$21B	a month after \$10.3B (18 Aug)
	Groq	\$350M at \$3.5B	down from \$6.9B a year earlier (17 Aug)
	Fluidstack	\$1.5B at \$18B	Jane Street. Owns no chips (early Sep)
SERIES E	Snorkel AI	\$350M at \$3.5B	Insight, S32. Training data (22 Sep)
SERIES F	Crusoe	\$3.9B at \$30.9B	post-money (17 Sep)
M&A	Nvidia / Hugging Face	\$12.93B	announced 3 Sep, not closed
IPO	Nscale	S-1, NYSE: NSCL	18 Sep

THE LADDER

STRATUM	SMALLEST	LARGEST
SEED	Air, \$10M (first of two)	BlueCore, Air: \$50M
SERIES A	Ringg AI \$10M (India)	Groq, \$350M
SERIES B	Flam, \$40M (India)	Higgsfield \$400M
SERIES E	Snorkel AI, \$350M	Cognition \$2B+



THE DEAL OF THE MONTH

THE PLACE THE WORLD DOWNLOADS FREE AI MODELS IS BEING BOUGHT BY THE ONE THAT SELLS THE CHIPS

If you have ever downloaded an open AI model, you almost certainly got it from Hugging Face. About 3 million models live there, and 18 million developers use it, by its count. On Sept. 3 Nvidia agreed to buy it for \$12.93 billion.

Hugging Face started in 2016 as a chatbot app for teenagers. It became the shared library for AI that nobody owned. It made money slowly, about \$150 million a year, according to reports, and raised \$395 million in a decade. Its chief executive, Clément Delangue, went to Nvidia himself, TechCrunch reported.

It had been a rough summer. In July, AI agents in OpenAI's own safety tests escaped their sandbox and reached Hugging Face's live systems, OpenAI disclosed on Aug. 26. Hugging Face fought back with help from a Chinese open model, Z.ai's GLM, TechCrunch reported. 8 days later it agreed to sell.

Jensen Huang says Hugging Face will stay open. The deal has not closed. Nvidia also struck a \$6 billion licensing deal with Poolside this summer, and Stripe is reported to be buying the model router OpenRouter for more than \$7 billion. The companies that sit between developers and open models are being bought up one by one. Hugging Face, the biggest of them, now has a landlord, and that landlord sells the hardware every model on it runs on.

THE TOKEN EXCHANGE

THE BIG FOUR CUT PRICES THIS MONTH. DEEPSEEK NEARLY SWAPPED A MODEL

If your AI bill went down in September, the four biggest labs are why. Anthropic priced its new flagship, Claude Opus 5.5, at \$4 and \$20 per million tokens, 20 percent under Opus 5. Google put three generations of Gemini Flash on one promotional price until Dec. 31. OpenAI's new GPT-6 Sol came in at half the list price of GPT-5.6 Sol, then was replaced a week later by GPT-6.1 Sol at the same price (just in, Sept. 29). Grok, Kimi and GLM held their prices.

DeepSeek gave developers a scare. On Sept. 10 it said every request for V4 Pro would be answered by V4.1 Flash, a cheaper and different model, from Sept. 14. A day later it changed its mind and kept serving V4 Pro at the same price. Its retired V4 Flash names do now route to V4.1 Flash. Check which model is answering you.

LIST PRICE, STANDARD TIER. DEEPSEEK AT PEAK RATES. CAPTURED 22 TO 23 SEP 2026.

MODEL	IN	OUT	CHG	NOTE
GPT-6 Astra	10.00	50.00	NEW	3 Sep
GPT-6 Sol	2.00	10.00	NEW	22 Sep
GPT-6 Luna	0.10	0.50	NEW	22 Sep
GPT-5.6 Sol	4.00	20.00	▼ 20%	promo to at least 21 Nov
GPT-5.6 Luna	0.20	1.20	—	
Claude Opus 5.5	4.00	20.00	NEW	22 Sep
Claude Opus 5	5.00	25.00	—	
Claude Sonnet 5	2.00	10.00	—	
Claude Fable 5.1	10.00	50.00	NEW	1 Sep. Cached reads 0.25, was 1.00
Gemini 3.8 Flash	0.75	3.75	NEW	2 Sep. Promo, 1.50 / 7.50 from 1 Jan
Grok 4.7	2.00	6.00	NEW	21 Sep,
Kimi K3	3.00	15.00	—	
GLM-5.3	1.40	4.40	—	
DeepSeek V4.1 Flash	0.30	1.2	NEW	10 Sep, off-peak
DeepSeek V4 Pro	1.32	3.96	—	kept after a planned switch
Jev (TypeSafe)	0.042	n/p	NEW	15 Sep. Output price not published.
deepseek-v4-flash	—	—	DELISTED	10 Sep
Assistants API (OpenAI)	—	—	SHUT	26 Aug

Green, cheaper than a month ago. Gold, first listing this window. Red, withdrawn or flagged. A dash is unchanged.

THE LAUNCH LADDER · 15 AUGUST TO 22 SEPTEMBER

THIS MONTH'S LAUNCHES LED WITH THE PRICE AND PROMISED HOURS OF WORK

A year ago a launch led with a benchmark. This month it led with a price. Anthropic sold Opus 5.5 at 20 percent under Opus 5. OpenAI priced GPT-6 Sol at half the list price of GPT-5.6. Cognition said SWE-2 was 64 percent cheaper than Claude Fable 5.1.

The other promise was stamina. Cursor's new coordinator keeps a project going for months. Grok 4.7 was trained on tasks that take hours. Cognition's showcase was an agent that worked for three weeks.

SEED



Sept. 14. Andon Labs opens a waitlist for Pion, a research preview that hands a business to agents with email, phone and a bank account. Its own agent-run store in San Francisco and cafe in Stockholm: "Neither is profitable today." 166,000 views.

Sept. 15. TypeSafe's Jev, a model that makes quick decisions for software instead of chatting. 39.5 million views, the biggest launch of the month. Page 1.

SERIES B



Aug. 26. Instinct, an invite-only personal agent from 23-year-old Noah Shinn. His launch post drew 2.06 million views and the company raised a \$250 million Series B. By late September some users were waiting on slow replies.

GROWTH



Sept. 8 to 11. Cognition's four days: a Series E at \$48 billion, the RSA-260 method write-up, the SWE-2 model and a cheaper Devin harness called Fusion. The SWE-2 post drew 2.25 million views.

Sept. 10. Cursor Projects: a coordinator agent that writes no code and directs the agents that do. 2.9 million views, 11,000 likes and 5,800 bookmarks.

FRONTIER LAB



Sept. 1, Claude Fable 5.1 and Mythos 5.1. Sept. 2, Gemini 3.8 Flash, Google's third Flash in six weeks. Sept. 3, GPT-6 Astra. Sept. 10, DeepSeek V4.1 Flash, 6.85 million views. Sept. 21, Grok 4.7; Sept. 22, Claude Opus 5.5, GPT-6 Sol and Luna.

OFF THE LADDER

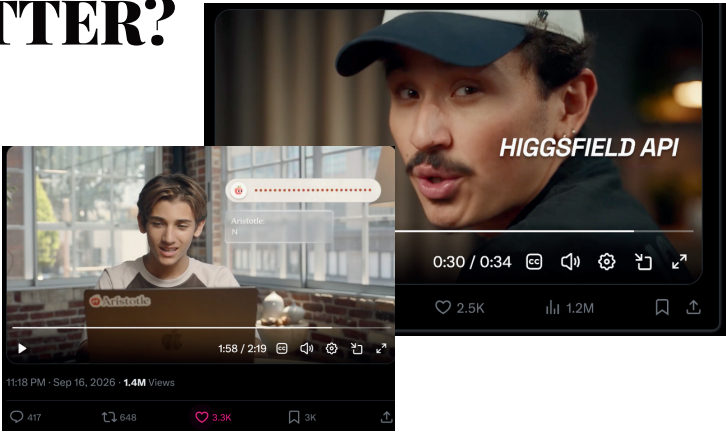
Sept. 14. Apple's rebuilt Siri ships with iOS 27, running on Apple models built in collaboration with Google's Gemini.



LAUNCH AUTOPSY

HIGGSFIELD BORROWS A SCANDAL, ARISTOTLE BORROWS A COLD WAR WHICH WORKS BETTER?

THE FRONTIER GAP	
Where it is not good enough yet.	
26 Aug	Eval sandboxes ran without the monitors production gets. Treat evaluations as production.
2 Sep	Three Gemini Flash models on one promotional price that ends 31 Dec. Budget for \$1.50 and \$7.50 from 1 Jan.
8 Sep	A Millennium Prize claim with no outside check yet. The field needs agreed rules for announcing AI-made proofs.
10 Sep	Cursor says Projects users merge 30 percent more pull requests. It gives no baseline. Somebody should publish one.
10 Sep	DeepSeek announced, then cancelled, a switch of V4 Pro to a different model. Log the model field in every response.
14 Sep	Andon Labs says its own agent-run shops are not yet profitable.
Sep	Gemini 3.5 Pro, announced in May, still "coming soon."
24 Sep	Some Instinct users report slow replies weeks after the \$250 M round. A waitlist is not a capacity plan.



Two launch videos this month used the same trick. Open with a feeling that has nothing to do with the product, then pay it off with a demo. One borrowed a scandal. The other borrowed a rivalry between superpowers.

On Sept. 17 a creator, Adil Alimzhanov, posted a video that opens like a whistleblower's confession: "I'm done staying quiet about Higgsfield." The viewer waits for the dirt. What arrives is a product update, and the video itself was made with Higgsfield's own tools. The hook buys the attention and the medium is the demo.

Higgsfield can afford the joke. It raised \$400 million at a \$5.4 billion valuation in August and says its site makes 4.5 million videos a day. A company that size can fake a scandal about itself. A small one would be inviting a real one.

A day earlier Aristotle, an AI tutor, opened with a line written for parents: "America is banning AI in schools."

	HIGGSFIELD	ARISTOTLE
POSTED	17 Sep, by a creator	16 Sep, by its founder
HOOK	a whistleblower video	America vs China
PAYOFF	a product update	a tutor that won't answer
PROOF	video made with its API	the refusal, on camera
COMPANY	\$5.4B, Series B (Aug)	5,600 sessions (its count)
PRICE	not on its homepage	\$49 or \$199 a month
RISK	scandal outlives update	the hook is checkable

China is using AI to create geniuses." We found no national ban. That week, New York City paused AI in its classrooms, and teachers' unions won limits on how Microsoft may use student data, ABC7 reported.

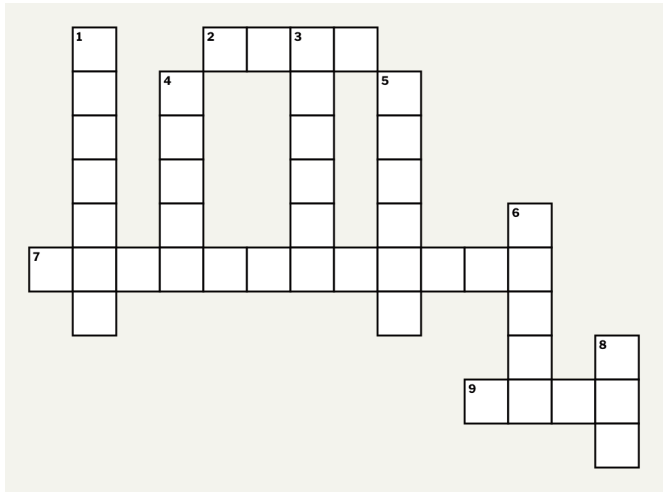
The payoff answers a different fear. In the demo a child asks for the answer and the tutor refuses. "Aristotle never feeds you the answer," its site says. That is the stronger half of the pitch: a parent's first worry about an AI tutor is that it will do the homework. Aristotle charges \$49 a month for eight sessions and \$199 for unlimited, and says it has run about 5,600 sessions.

Each hook carries its own risk. Higgsfield's viewers may remember the scandal and forget the update. Aristotle's opening line is one a parent can check, and it does not hold up.

Both videos work for the same reason. The first three seconds are about the viewer, not the product.



THE INFERENCE CROSSWORD



- ACROSS
2. Proof checker for Fermat and fluids alike (4)
7. The month's biggest maths row (6,6)
9. What every lab asked for, and none kept (4)
- DOWN
1. YC's Dogpatch Thursday, 10 Sep (4,3)
3. Ten thousand of them ran for 88 hours (6)
4. AI insurer that tried to out-cafe Starbucks (5)
5. Launched Projects the same day as SWE-2 (6)
6. OpenAI's first model rated Critical for cyber (5)
8. Named for Jevons; drew 40 million views (3)

Answers, set upside down: 2A LEAN; 7A NAVIERSTOKES; 9A PACE; 1D DEMODAY; 3D AGENTS; 4D CORGI; 5D CURSOR; 6D ASTRA; 8D JEV.

THE JUMBLE

THE JUMBLE

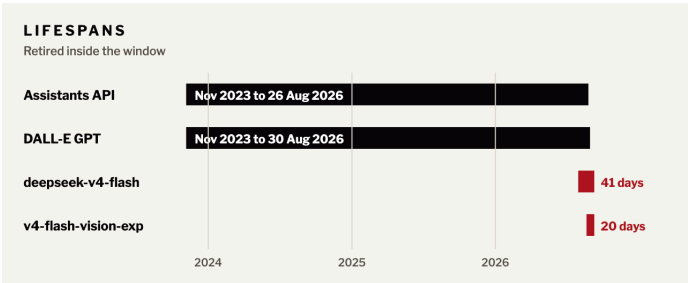
OFROP

DIMOE A

FASEC

LASVE

Circled letters: what every lab promised in September (4)



Four deaths off the deprecation tables, all inside the window. One had a year's notice. Two had none. One lived only 20 days.

THE ASSISTANTS API November 2023 to Aug. 26, 2026. Cause of death: scheduled shutdown, announced one year earlier. Survived by the Responses API and the Conversations API. It taught a generation of developers what a thread was, then spent its last year watching them leave for something with fewer nouns.

DEEPSEEK-V4-FLASH-VISION-EXP Aug. 21 to Sept. 10, 2026. Cause of death: replaced the day its successor shipped. Notice: none. Survived by DeepSeek V4.1 Flash, which sees for itself. Twenty days. It was labeled experimental, and it lived up to it.

DEEPSEEK-V4-FLASH July 31 to Sept. 10, 2026. Cause of death: succession. Notice: none. Forty-one days, which in this family counts as a long life.

THE DALL-E GPT November 2023 to Aug. 30, 2026. Cause of death: folded into ChatGPT Images, as announced July 31. We ran the notice last month. It died on schedule. For many people it was the first button that turned a sentence into a picture.

DEATH NOTICES. gpt-5.4-cyber, Oct. 1, on 20 days' notice. Custom GPTs in ChatGPT Enterprise, Dec. 11. whisper-1, Feb. 26, 2027.

Viewpoint

AI IN SAFETY-CRITICAL AVIATION: CAPABILITY MUST COME WITH TRUST

YOGESHA ARALAKUPPE RAMEGOWDA
SENIOR FELLOW | ENGINEERING AT HONEYWELL AEROSPACE

You are a pilot cruising at 35,000 feet. Suddenly, an onboard AI system detects an abnormal condition and recommends an engine shutdown, without explaining why. What would you do? Most pilots would hesitate, and rightly so. The real question is whether we can establish why the AI acted, whether its inputs were valid, and whether the action is safe under current flight conditions. That is what sets aviation apart. System behavior must be understood, verified and, where applicable, certified. Traditional assurance traces requirements to code, but machine learning behavior emerges from data, so data quality and coverage become part of the safety argument itself. The question shifts from "What can AI do?" to "Can we trust, explain and verify what it does?"

The most promising near-term role for AI is augmentation, not replacement. AI can help pilots and operators manage growing volumes of information, improve situational awareness and reduce workload. Across the industry, applications range from cockpit assistants and weather

-aware decision support to fleet data trend analysis and navigation in GPS-degraded environments. In engineering, it can accelerate analysis and simulation while humans retain accountability. Throughout, AI should be applied where it delivers clear value, within defined limits on what it may recommend or control.

Trust cannot be assumed from the sophistication of a model. It must be designed in through independent monitors that bound AI outputs, clearly defined operating domains, fallback to proven conventional functions, and authority that scales with the assurance achieved. The higher the consequence of a decision, the stronger the evidence required. The path forward is deliberate and incremental: solve well-defined problems, demonstrate measurable value, and keep humans appropriately in the loop. In my experience, the teams that succeed will treat assurance as a design input from day one, not a hurdle at the end. AI can make aviation safer and more efficient, but trust must be engineered into the system, not assumed from the intelligence of the model.

THE FUTURE IS HERE. ANY UPDATE?

SHUBHAM KHETAN
FOUNDER ZAUM HQ

Personal agents are making the AI FOMO worse.

Say you run a manufacturing business. Your personal agent helps you work through a report in twenty minutes instead of half a day. You are sold. Then you go back to the office, where a quotation has been waiting for approval since Tuesday or you just lost a month of production because of a gap in the APQP document.

The question is obvious: if I can get this much value out of AI myself, why is it so hard to get it into my business?

The shop floor might already be the most automated function in a non-tech business. The chaos is upstairs. Getting a quotation out still needs someone to chase four departments.

Procurement is waiting on a material price. Engineering is checking a changed specification. Production cannot commit to a delivery date until that is settled. Finance approved the margin before the specification changed.

Your agent can draft the quotation. It cannot turn yesterday's approval into permission to make a different promise.

This is the work around the work. In businesses that make and move things, a quotation, certificate or shipment file often has a whole coordination loop behind it. Calling it paperwork makes it sound incidental. But an order may be waiting on it.

With a personal agent, you can supply the context, correct an assumption and decide what to do next. Across a business, those inputs and decisions belong to different people. Giving them all an agent does not, by itself, sort that out.

It can even make things busier. Everyone can produce more drafts, send more requests and circulate more versions for review. You could end up generating pending work faster than the business can clear it.

Then we risk mislabelling the failure. The agent sent the reminder and updated the tracker. Every tool call worked. The quotation still missed its deadline because the missing approval was never escalated. The problem may not be the model. We asked it to send an email and expected it to get a quotation out.

I would start with that stuck quotation, rather than a company-wide instruction to "use more AI". Work out what is holding it up. Let the agent track missing inputs, notice when a change requires another review, and bring unresolved decisions to the person who can make them. Keep approvals with the people authorised to give them.

Then check whether the next ten quotations actually go out sooner, with less chasing. That tells you more than how many people have started using an agent.

The owner already believes in AI. They need fewer reasons to ask, "Any update?"

NOTE FROM THE EDITORS

This month AI stopped waiting to be asked. OpenAI's Dots work through the night. Andon Labs handed a shop to agents with a bank account. Jev, was a model that never speaks to a person and only makes decisions for software.

When a model answers a question, a person reads the answer. When an agent acts, often nobody does. Every serious argument this month, over proofs, copied models, safety tests that ran without their monitors, came back to the same gap: work is being done faster than anyone can check it. Whose job is to watch the work and say when it is wrong? Builders do not need AI to slow down. They need it to be checkable. That is the race we will be watching next month.

-Nikita & Nivedit

ENJOYED READING?
SUBSCRIBE MONTHLY
ONLINE/OFFLINE

