

xAI releases Grok 4 and a SuperGrok Heavy tier for Grok 4 Heavy

Source copy kept by wheresthe.ai for the story "xAI releases Grok 4 and a SuperGrok Heavy tier for Grok 4 Heavy" (<https://wheresthe.ai/d/xai-grok-4-release-2025-07/>).

Source: <https://x.ai/news/grok-4>

Text saved 18 Oct 2026 (the page as published).

SHA-256 of the saved text: 1c37e181c430903ef571d69be62d5be0500a5fe83d2345589094132e070400d0

Grok 4 | SpaceXSI Back to news Jul 9, 2025 Grok 4 Grok 4 is the most intelligent model in the world. It includes native tool use and real-time search integration, and is available now to SuperGrok and Premium+ subscribers, as well as through the xAI API. We are also introducing a new SuperGrok Heavy tier with access to Grok 4 Heavy - the most powerful version of Grok 4. Try SuperGrok Access the API

Scaling Up Reinforcement Learning With Grok 3, we scaled next-token prediction pretraining to unprecedented levels, resulting in a model with unparalleled world knowledge and performance. We also introduced Grok 3 Reasoning, which was trained using reinforcement learning to think longer about problems and solve them with increased accuracy. During our work on Grok 3 Reasoning, we noticed scaling trends that suggested it would be possible to scale up our reinforcement learning training significantly. For Grok 4, we utilized Colossus, our 200,000 GPU cluster, to run reinforcement learning training that refines Grok's reasoning abilities at pretraining scale. This was made possible with innovations throughout the stack, including new infrastructure and algorithmic work that increased the compute efficiency of our training by 6x, as well as a massive data collection effort, where we significantly expanded our verifiable training data from primarily math and coding data to many more domains. The resulting training run saw smooth performance gains while training on over an order of magnitude more compute than had been used previously. Humanity's Last Exam Deep expert-level benchmark at the frontier of human knowledge

State of the art Full set (April 3, 2025) with Python and Internet tools

Performance over training Text-only subset with Python and Internet tools

Pass@1 Accuracy (%) 60

Compute

No tool

With tool

Test time compute TTC

Native Tool Use Grok 4 was trained with reinforcement learning to use tools. This allows Grok to augment its thinking with tools like a code interpreter and web browsing in situations that are usually challenging for large language models. When searching for real-time information or answering difficult research questions, Grok 4 chooses its own search queries, finding knowledge from across the web and diving as deeply as it needs to craft a high-quality response. We also trained Grok to use powerful tools to find information from deep within X. Grok can use advanced keyword and semantic search tools and even view media to improve the quality of its answers. X Search Web Search Research

I remember this popular post from a few days ago about this crazy word puzzle which had something to do with legs. Can you help me find it?

Show entire trace Thought for 1 minute

Investigating puzzle post The request is for a popular post from early July 2025 about a crazy word puzzle involving legs. Let's try searching for relevant X posts from July 1 to July 9, 2025. Maybe searching for 'viral word puzzle about legs' could help find recent posts. Searching for viral word puzzle about legs

🔍 @iquantified 🔍 @skeletonNaughty 🔍 @AZ_NYsles + 3 others

Searching for (word puzzle) legs min_faves:100 since:2025-07-01 until:2025-07-09 filter:has_engagement

✂ @LegsEmporium ✂ @chelsbrina13

Posts from the search show a puzzle about words ending with homophones for leg parts like toe, calf, knee, shin. Many posts mention 'toe, calf, knee, shin' as the purple category in Connections puzzle #756. Exploring puzzle details Posts around July 7-8, 2025, suggest the viral post is about the NYT Connections puzzle's absurd category. Let's search for 'NYT Connections puzzle #756 answers'. Searching for NYT Connections puzzle #756 answers

falsenytimes.com falsemashable.com falsecnet.com + 2 others

For July 6, 2025, puzzle #756's purple category was words ending with homophones of leg parts. The 'crazy' part is the stretch, like 'shun' sounding like 'shin' in some accents. Searching for connections (toe calf knee shin) since:2025-07-06 until:2025-07-09 min_faves:100

✂ @zaqtff

Searching for nyt connections crazy leg puzzle homophones

✂ @DreamWeaver5166 ✂ @iquantified ✂ @VKIker + 3 others

A post with 660 likes mentions 'toe calf knee shin they're homophones,' which could be the popular one. Now, let's search X for posts with 'homophones' and 'parts of the leg' from July 6 to 9, 2025, requiring at least 100 favorites. Searching for homophones "parts of the leg" since:2025-07-06 until:2025-07-09 min_faves:100

✂ @ThaTurnerIV

Searching for connections legs OR leg since:2025-07-06 until:2025-07-09 min_faves:500

✂ @IsOutOf0

Analyzing search results A new search found a reply post with 5052 likes, quoting a meme about the puzzle's absurdity. Another search revealed a post with 89123 likes, likely the viral one, quoting a complaint about the puzzle.

Based on your description, it sounds like you're referring to a viral X post from July 7, 2025, complaining about a particularly frustrating category in the New York Times' daily Connections word puzzle (game #756). The puzzle involves grouping 16 words into four themed categories, and one of them—the purple (hardest) group—was "ending with homophones for parts of the leg." The words were crypto (sounds like it ends with "toe"), decaf (ends with "calf"), Disney (ends with "knee"), and Prussian (ends with "shin"). This stretched connection sparked a lot of outrage and memes online for being overly obscure and accent-dependent. The specific popular post you might be thinking of shows a screenshot of the puzzle grid with the caption "This game is a psyop to give me schizophrenia." It was posted by @LinkofSunshine and quickly went viral, racking up over 21,000 likes and hundreds of replies, quotes, and reposts within days. Many people in the replies echoed the frustration, calling the leg-related category "ridiculous" or "impossible without guessing." If this isn't the exact one you meant, let me know more details (like the username or exact phrasing), and I can dig deeper!

Grok 4 Heavy We have made further progress on parallel test-time compute, which allows Grok to consider multiple hypotheses at once. We call this model Grok 4 Heavy, and it sets a new standard for performance and reliability. Grok 4 Heavy saturates most academic benchmarks and is the first model to score 50% on Humanity's Last Exam, a benchmark "designed to be the final closed-ended academic benchmark of its kind."

Grok 4 Heavy

Processing ~ 10 min left

Agent 1 ~ 10 min left

Agent 2 ~ 10 min left

Agent 3 ~ 10 min left

Thought for 10 minutes

Frontier Intelligence Grok 4 represents a leap in frontier intelligence, setting a new state-of-the-art for closed models on ARC-AGI V2 with 15.9% (nearly double Opus's ~8.6%, +8pp over previous high). On the agentic

Vending-Bench, it dominates with \$4694.15 net worth and 4569 units sold (averages across 5 runs), vastly outpacing Claude Opus 4 (\$2077.41, 1412 units), humans (\$844.05, 344 units), and others. Grok 4 Heavy leads USAMO'25 with 61.9%, and is the first to score 50.7% on Humanity's Last Exam (text-only subset), demonstrating unparalleled capabilities in complex reasoning through scaled reinforcement learning and native tool use. GPQA Science

LiveCodeBench (Jan - May) Competitive Coding

USAMO 2025 Olympiad Math Proofs

HMMT 2025 Competitive Math

AIME'25 Competition Math

ARC-AGI-2 Abstraction and Reasoning

Grok 4 API The Grok 4 API empowers developers with frontier-level multimodal understanding, a 256,000 context window, and advanced reasoning capabilities to tackle complex tasks across text and vision. It integrates real-time data search across X, the web, and various news sources via our newly launched live search API, enabling up-to-date, accurate responses powered by native tool use. With enterprise-grade security and compliance—including SOC 2 Type 2, GDPR, and CCPA certifications—the API ensures robust protection for sensitive applications. Grok 4 is coming soon to our hyperscaler partners, making it easier for enterprises to deploy at scale for innovative AI solutions. Grok 4 Voice Mode Speak with Grok in our upgraded Voice Mode, which features enhanced realism, responsiveness, and intelligence. We introduce a serene, brand-new voice and redesign conversations to make them even more natural.

And now, Grok can see what you see! Point your camera, speak right away, and Grok pulls live insights, analyzing your scene and responding to you in real-time from within the voice chat experience. We are proud to present this model trained in-house, with our state-of-the-art reinforcement learning framework and speech compression techniques.

Enable video during your voice chat and Grok will look at what it sees when talking to you. What's Next xAI will continue scaling reinforcement learning to unprecedented levels, building on Grok 4's advancements to push the boundaries of artificial intelligence. We plan to expand the scope from verifiable rewards in controlled domains to tackling complex real-world problems, where models can learn and adapt in dynamic environments. Multimodal capabilities will see ongoing improvements, integrating vision, audio, and beyond for more intuitive interactions. Overall, our focus remains on making models smarter, faster, and more efficient, as we drive toward systems that truly understand and assist humanity in profound ways.