

xAI releases Grok 4.6 for long-running agents

Source copy kept by wheresthe.ai for the story "xAI releases Grok 4.6 for long-running agents" (<https://wheresthe.ai/d/xai-grok-4-6-release-2026-08/>).

Source: <https://x.ai/news/grok-4-6>

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

SHA-256 of the saved text: d6ceb5394e384e4f53f8d15f67026431304529855e9e2ed47f93f1da2a8b2dbc

Introducing Grok 4.6 | SpaceXSI Back to news Aug 12, 2026 Introducing Grok 4.6 Grok 4.6 builds on Grok 4.5 with a particular focus on long-running agents and more ambitious interactive and visual work. Try for free Start building

Today we are releasing Grok 4.6 . Grok 4.6 builds on Grok 4.5 with a particular focus on long-running agents and more ambitious interactive and visual work. It stays with complex tasks across many steps, whether researching a topic, analyzing information, working across a codebase, or turning an idea into a polished application or work artifact.

0:00 / 0:00

Grok 4.6 achieves frontier intelligence across several agentic coding and knowledge work benchmarks. It matches GPT-5.6 Sol on the Artificial Analysis Intelligence Index, which is a composite score of nine benchmarks. AA Intelligence GDPVal-AA DeepSWE 1.1 CursorBench 3.2 FrontierCode 1.1

Competitor figures are drawn from the respective developers' published system cards or benchmark leaderboards

Benchmark bar charts comparing Grok 4.6 with other leading models across AA Intelligence, GDPVal-AA, DeepSWE 1.1, CursorBench 3.2, and FrontierCode 1.1. Competitor figures are drawn from the respective developers' published system cards or benchmark leaderboards. Grok 4.6 is available today in Cursor and Grok Build . We're offering 2x included usage inside Grok Build and Cursor for the first week so you can start trying 4.6 immediately. Training Grok 4.6 Grok 4.6 underwent a longer supplemental training run than Grok 4.5, with curated model-generated data for reasoning and advanced technical concepts, high-quality engineering data, and an improved optimizer and training recipe. This produced a stronger foundation for the SFT and RL stages that followed. We then used Grok 4.5 to regenerate the SFT trajectories across reasoning efforts, agent harnesses, and domains such as STEM, software engineering, and knowledge work, and filtered out problematic traces with model-based checks. The resulting SFT checkpoint shows strong performance and improved behavior. Grok 4.6 is trained on a wide range of agentic RL tasks, including knowledge work, general coding, and domain-specific environments for kernel optimization, web development, computer-aided design, and more. Turning ambitious ideas into working projects We tested Grok 4.6 on projects designed to stretch its range and ability to sustain work over many steps. We found the model is especially strong at turning a broad product idea into a working first version. It can research unfamiliar domains, structure the application, implement the core interactions, and continue refining the result through several rounds of feedback. On longer trajectories, we also started to see more self-testing and verification, with the model checking its own work before moving on. Grok 4.6 produces stronger first passes on visual and interactive projects than we typically saw with Grok 4.5. Given a concrete product idea, it is able to establish structure and visual language for an application in one pass. This has made it especially useful for projects where the fastest route to a good result was to begin with something substantial and then iterate in the loop. Safety and capabilities Grok 4.6's safeguards have been improved and calibrated in line with the model's capabilities. Our safety stack is designed to maximize utility and security across legitimate use cases, allowing Grok 4.6 to be helpful and safe in domains such as vulnerability patching, accelerating the engineering design cycle, and augmenting AI research. Our safeguard evaluation work reflects Grok 4.6's expanded capabilities, with our widest-ever suite of pre-deployment testing for capabilities and safeguard calibration, as well as extensive post-deployment and third-party testing. Evals

Grok 4.6 High Grok 4.5 High GPT-5.6 Sol Max Fable 5 Max

AA Intelligence Index

61 56 61

62

GDPVal-AA v2

1753 1526 1728 1741

CursorBench v3.2

69.9% 66.7% 67.2%

70.5%

DeepSWE v1.1

65.9% 54%

73% 70%

FrontierCode v1.1 (Extended)

61.3% 56.6% 60.6%

63.6%

APEX-Agents

57.5% 47.1% 56.7%

59.2%

Terminal-Bench v3.0

26% 15.7%

34.6% 34.1%

APEX-SWE

56.4% 53.6% —

58.8%

AA-Briefcase

1577 1313 1502 1574

Harvey LAB (Vals)

15.8% 12.9% 2.5% 11.3%

Best score per evaluation in bold. Third-party model scores are the best of self-reported or publicly available results.

Get started with Grok 4.6 Grok 4.6 is available today in Cursor and Grok Build . It's also available in the API and other partners like OpenRouter, Vercel, and Cloudflare. Pricing starts at \$2 per million input tokens and \$6 per million output tokens. Additionally, there is a fast variant which is twice the price. We're offering 2x included usage inside Grok Build and Cursor for the first week so you can start trying 4.6 immediately.

Console Create an API key

docs.x.ai Read the docs

Try it in Grok Build for free Get started today at x.ai/build . \$ curl -fsSL <https://x.ai/cli/install.sh> | bash