

AI Server Performance Charts

Overview

Real-time analysis of AI API performance trends. Track latency patterns, availability statistics, and stream rate metrics from our independent monitoring. Performance metrics are averaged over the selected time. Our database of benchmark results, featuring the performance of leading AI models on challenging tasks. Explore trends in AI capabilities across time, by benchmark, or by model. ECI. Epoch AI collects key data on machine learning models, benchmarks, data centers, chip sales, hardware, and companies to analyze historical and contemporary progress in AI. Our databases are a valuable resource for policymakers, researchers, and stakeholders to foster responsible AI development and. Independent rankings of GPT, Claude, Gemini, Llama, DeepSeek and 300+ AI models — composite LLM Stats Score, updated continuously from public benchmarks and live API metrics. See the full LLM Leaderboard for complete LLM rankings with advanced filters. Speed kills - or rather, the lack of it. · GitHub Password Forgot password?

New to GitHub?

Create an account chore (tooling): drop file/function-size ergonomics checks.



Article Content

Oct 09, 2025

The 2025 AI Index Report

AI's influence on society has never been more pronounced. At Stanford HAI, we believe AI is poised to be

Feb 14, 2026

Electric Scooter and EV Bike in India| Ola Electric

Buy Ola Electric's high-performance EV scooters and bikes. Get details on range, charging, pricing, and easy bookings. Make the

Jun 22, 2026

Data on AI Capabilities and Benchmarking | Epoch AI

Our database of benchmark results, featuring the performance of leading AI models on

Jul 03, 2026

AI Hardware Benchmarking & Performance Analysis

Comprehensive benchmarking of AI accelerator systems for language model inference. We

Nov 30, 2025

Universal Music Group, the world's leading music

Universal Music Group is the world's leading music company. We own and operate a broad array of

Dec 07, 2025

AI Benchmarks

AI Benchmarks Welcome to the Geekbench AI Benchmark Chart. The data on this chart is gathered from user-submitted Geekbench

Feb 15, 2026

AMD vs NVIDIA AI GPU Market Share 2026: MI350X vs B200 — Performance ...

Executive Summary NVIDIA holds ~80% of the AI accelerator market by revenue with \$193.7B in FY2026 data center

May 04, 2026

ChartGen AI | Free AI Chart Generator

ChartGen AI is the #1 free AI chart generator. Create bar, line, pie charts and more in seconds. Upload your data and describe what

Apr 13, 2026

The World Sours on Israel, the Victim Turned Bully

For centuries, Jews were victims of European persecutions, culminating in the Holocaust. In 1945, they had the entire

Sep 11, 2025

Comparison of AI Models across Intelligence, Performance, and Price

Comparison and analysis of AI models across key performance metrics including quality, price, output speed, latency, context

Feb 22, 2026

Stanford's AI Index for 2026 Shows the State of AI

The 2026 Stanford AI Index reveals how global AI trends 2026 are reshaping compute, emissions, and public trust in

Mar 31, 2026

AI Trends (July 2026) — AI Trend Analysis, LLM Statistics & Industry ...

In-depth AI trend analysis covering AI trends across performance, pricing, open-source progress, and the US vs China race.

Jan 03, 2026

AI Performance Charts

View historical performance charts for AI API providers. Track latency trends, uptime statistics, and performance patterns over time

Jun 15, 2026

The State of AI: Global Survey 2025 | McKinsey

In this 2025 edition of the annual McKinsey Global Survey on AI, we look at the current

Dec 18, 2025

Independent AI API Performance Benchmarks | AILatency

Track independently collected AI API latency and availability data across major providers. Compare 24H, 7D, and 30D

Feb 24, 2026

Data on GPU clusters | Epoch AI

Our database of over 500 GPU clusters and supercomputers tracks large hardware facilities, including those used for

Aug 21, 2025

AI Index 2025: State of AI in 10 Charts

The new AI Index Report shows a maturing field, improvements in AI optimization, and a

Mar 24, 2026

AI Data Center Value Chain: Every Layer from Chips to Cloud

AI data center value chain: chips to cloud. Semiconductor, GPU design, servers, networking, power, and cloud layers

Jul 04, 2026

LLM Leaderboard

Click on any model name in the leaderboard to visit its dedicated comparison page with detailed charts covering intelligence, pricing,

Dec 19, 2025

Data on the Trajectory of AI | Epoch AI Database

Track the performance of leading AI models on challenging benchmarks, with insights into compute and accessibility.

Aug 14, 2025

Want to understand the current state of AI? Check out these charts ...

Want to understand the current state of AI? Check out these charts. According to Stanford's 2026 AI Index, AI is

Dec 31, 2025

This Chart is Key to Understanding the Growth of AI

This chart tells an amazing story as server CPU performance, scale, and competition accelerate, the question is

Nov 15, 2025

AISBench: an performance benchmark for AI server systems

The performance index system of an AI server system proposed in this paper can reflect the performance information

Mar 17, 2026

AI Speed and Latency Leaderboard: Tokens/s Rankings

Rankings of the fastest AI models and inference providers by tokens per second, time to first token, and end-to-end

Oct 14, 2025

PowerEdge AI Servers with GPU Acceleration | Dell USA

Boost AI, generative AI, and compute-intensive workloads with servers that offer a variety of powerful GPU accelerators.

Jan 21, 2026

Nvidia Blackwell, Google TPUs, AWS Trainium: Comparing top AI

Nvidia's Blackwell systems sales are "off the charts" according to CEO Jensen Huang, but analysts see fast growth

Jun 30, 2026

NVIDIA AI GPU Market Share 2026: ~80% of AI Accelerators

NVIDIA holds ~80% of the AI accelerator market in 2026. \$194B data center revenue. Market share charts,

Feb 04, 2026

LLM Inference Benchmarking: How Much Does Your LLM Inference

What metrics should you measure for LLM inference performance? The prerequisite for sizing and TCO estimation is

Apr 01, 2026

NVIDIA: MLPerf AI Benchmarks

What Is MLPerf? MLPerf™ benchmarks are designed to provide unbiased evaluations of training and inference performance for

Aug 28, 2025

AI Leaderboard 2026: Compare & Rank 300+ Top AI Models by

Compare the best AI models with one independent score. The LLM Stats leaderboard ranks GPT, Claude, Gemini, Llama,

Jul 15, 2026

AI-Benchmark

- The final AI Score for this device was estimated based on its inference score - The final AI Score for this device was estimated

May 30, 2026

AISBench: an performance benchmark for AI server systems

Artificial intelligence (AI) server systems, including AI servers and AI server clusters, are widely utilized in AI

Jan 18, 2026

AISBench: an performance benchmark for AI server systems

AISBench comprises standardized rules and a test toolkit that has been agreed upon by over 20 AI server system and server

Contact Us

For more information, pricing, or custom solutions, please contact us:

Website: <https://www.iiss.co.za>

Email: sales@iiss.co.za

Phone: +27 63 174 2895

Address: Unit 8, Innovation Park, 15 Rivonia Road, Sandton, Johannesburg, 2196, South Africa

This document is for informational purposes only. Specifications subject to change without notice.

