

China s first AI inference server



Overview

Chinese scientists have developed a compact, energy-efficient AI server called BI Explorer 1 (BIE-1), which mimics human brain function. This mini-fridge-sized device boasts impressive computational power while consuming significantly less energy than traditional supercomputers. The BI Explorer computing system, or BIE-1, was revealed by members of the TeleChat3 series – China Telecom's TeleAI released the first large-scale Mixture-of-Experts (MoE) models trained entirely on domestically designed semiconductors. Domestic hardware stack – Training was conducted exclusively on Huawei's Ascend 910B AI chips and the open-source MindSpore framework. It became the first product to realize the single-machine inference service of the 671B DeepSeek large model on a domestic GPU, providing the industry with an integrated large-model intelligent application solution. JD Cloud also released the DeepSeek large model all-in-one machine, which supports. Guangdong Institute of Intelligence Science and Technology (GDIIST) has unveiled the BI Explorer (BIE-1), a mini-fridge-sized AI server it has said can rival a supercomputer while using one-tenth of the power, with specs including 1,152 CPU cores and 4. Chinese scientists. As Baidu's AI chip unit targets a \$50 billion Hong Kong IPO — 40% above its parent's entire market cap — a landmark cross-rival procurement deal reveals that China's internet giants can no longer afford to build everything alone.

Article Content

Azure updates | Microsoft Azure

Subscribe to Microsoft Azure today for service updates, all in one place. Check out the new Cloud Platform roadmap to see our latest product plans.

Report: China's DeepSeek follows OpenAI in developing its own

DeepSeek Ltd., one of China's most visible artificial intelligence companies, is pushing to design its own, in-house silicon aimed at inference workloads, according to a report by Reuters today ...

China unveils "mini fridge" AI server that uses 90% less power than ...

The device packs the capabilities of a supercomputer – including training and inference speeds that rival traditional computing clusters – into a refrigerator-sized unit with the help of an

Gartner Business Insights, Strategies & Trends For

Gain strategic business insights on cross-functional topics, and learn how to apply them to your function and role to drive stronger performance and innovation.

GenAI fuels China's \$223B compute surge

Global AI server market to hit \$158.7B in 2025, up from \$125.1B in 2024, with generative AI servers accounting for nearly 30% of total spend. China's AI compute capacity to reach 2,781.9

eeNews Analog ...

TDK Corporation has developed a prototype analog reservoir AI chip capable of real-time learning, in collaboration

VentureBeat | Transformative tech coverage that matters

VentureBeat delivers news, analysis, and insights on AI, data, and security—helping business leaders stay ahead in the rapidly evolving tech landscape.

Domestic DeepSeek all-in-one machine: new products burst out, new ...

It became the first product to realize the single-machine inference service of the 671B DeepSeek large model on a domestic GPU, providing the industry with an integrated large-model

50 Top AI Funded Startups (June 2026)

Top 50 AI Funded Startups June 2026 Complete rankings of the most funded artificial intelligence companies. Anthropic overtook OpenAI in private valuation with a \$65B Series H at

Tech News | Today's Latest Technology News | Reuters

Find latest technology news from every corner of the globe at Reuters , your online source for breaking international news coverage.

Tencent Buys Baidu Kunlunxin AI Chips — \$50B IPO Signal

Tencent's cross-rival procurement of Baidu's Kunlunxin AI chips, paired with a \$50B Hong Kong IPO targeting, marks the end of China's big-tech closed-ecosystem era and a new

China builds brain-mimicking AI server the size of a mini-fridge ...

Chinese scientists have unveiled what they call the "world's first brain-inspired computing entity," an AI "supercomputer" called the BIE-1, inspired by the operation of the human brain.

NVIDIA Kicks Off the Next Generation of AI With Rubin

NVIDIA today kickstarted the next generation of AI with the launch of the NVIDIA Rubin platform, comprising six new chips designed to deliver one

China unveils "brain-like" AI server that uses 90% less ...

China's Guangdong Institute of Intelligence Science and Technology (GDIIST) has revealed a compact, "brain-like" AI server that it has said can deliver supercomputer-class

M365 Cross-Cloud and Cross-Tenant Collaboration Scenarios

As more of our customers onboard onto our government clouds (GCC, GCC High and DoD) they also need to be able to collaborate across those boundaries. Many cross-cloud scenarios

China's Telecom Operators Take the Lead in AI

Today's column examines China's AI infrastructure buildout: RMB 51 billion in inference servers, 222 procurement projects, and a shift in what kind of

Technology news and analysis | The Register

The Register is an iconic enterprise technology news website read by millions of IT pros around the world

Wiley Online Library

Hier sollte eine Beschreibung angezeigt werden, diese Seite lässt dies jedoch nicht zu.

Sage Journals: Your gateway to world-class journal research

Hier sollte eine Beschreibung angezeigt werden, diese Seite lässt dies jedoch nicht zu.

News Archive

As AI moves from model development to production inference, compute demand is accelerating and shifting toward continuously operating AI

China Unveils Groundbreaking Brain-Inspired AI Server:

Chinese scientists have developed a compact, energy-efficient AI server called BI Explorer 1 (BIE-1), which mimics human brain function. This mini

China unveils brain-like AI server that cuts power use by 90%

As AI becomes central to business and research, scientists face a pressing challenge – how to make it vastly more efficient. In China, one research team claims to have taken a major step

China Telecom built AI models with home-grown hardware

China Telecom's AI research arm, TeleAI, has open-sourced the TeleChat3 series of large language models, which are China's first large-scale

vocab.txt · nomic-ai/nomic-embed-text-v1.5 at refs/pr/55

We're on a journey to advance and democratize artificial intelligence through open source and open science.

Japan Targets Sovereign AI Model and 10 Million Robots

2 The government targets 10 million AI-equipped robots across 18 sectors by 2040, creating a sustained engineering demand curve for edge inference, domain-specific datasets, and

Energy and AI – Analysis

The development and uptake of artificial intelligence (AI) has accelerated in recent years – elevating the question of what widespread

Changes with Storage on Cloud and Storage Classification in VMM

First published on TECHNET on May 19, 2014 Storage Classification was introduced in System Center 2012 Virtual Machine Manager (VMM 2012) to provide the...

Contact Us

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

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

Email: info@introspec.co.za

Phone: +27 73 849 2156

Address: 25 Riebeek Street, Cape Town, 8001, South Africa

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

