MaxBits · Daily Tech Watch

High-quality technology news from around the world.

Daily brief · 2026-02-13

Open Weekly view (local)   Weekly = articoli selezionati con "Add to Weekly", salvati solo nel tuo browser.

Last 7 daily reports

3 deep-dives you should really read

OpenAI deploys Cerebras chips for 'near-instant' code generation in first major move beyond Nvidia

VentureBeat – AI · Topic: **AI/Cloud/Quantum**

- **What it is:** OpenAI has launched GPT-5.3-Codex-Spark, a specialized coding model engineered for near-instantaneous code generation using Cerebras wafer-scale processors for inference workloads.
- **Who:** OpenAI, a leading AI developer, partners with Cerebras Systems, a specialized chipmaker, diversifying its hardware supply beyond dominant incumbent Nvidia for specific AI inference tasks.
- **What it does:** The new model, running on Cerebras chips, delivers near-instantaneous code generation for developers, significantly improving real-time coding workflows by prioritizing speed over peak capability for specific tasks.
- **Why it matters:** This move signifies OpenAI's strategic diversification of chip suppliers beyond Nvidia, addressing latency-sensitive inference workloads and potential supply chain vulnerabilities amid a strained relationship with its traditional provider.
- **Strategic view:** This partnership signals AI providers' increasing pursuit of specialized, low-latency hardware for inference, diversifying supply chains away from Nvidia to optimize user experience and create new interaction patterns for real-time AI applications.

☐ Add to Weekly

Google Chrome ships WebMCP in early preview, turning every website into a structured tool for AI agents

VentureBeat – AI · Topic: **AI/Cloud/Quantum**

- **What it is:** WebMCP is a new proposed web standard, launched by Google Chrome as an early preview, enabling websites to expose structured, callable tools directly to AI agents via a browser API.
- **Who:** Google Chrome, in collaboration with Microsoft and W3C, is shipping WebMCP, a new protocol shaping how AI agents interact with web content.
- **What it does:** WebMCP allows websites to expose structured functions directly to AI agents, streamlining interactions and enabling collaborative browsing, contrasting with previous inefficient scraping methods.
- **Why it matters:** It significantly reduces the cost and improves the reliability of AI agent interactions with websites, moving beyond fragile scraping to direct, structured function calls.
- **Strategic view:** This standard fundamentally transforms web interaction by integrating AI agents directly into the browser, fostering collaborative human-AI experiences and driving efficiency in digital product development.

☐ Add to Weekly

AI inference costs dropped up to 10x on Nvidia's Blackwell — but hardware is only half the equation

VentureBeat – AI · Topic: **AI/Cloud/Quantum**

- **What it is:** AI inference costs have dropped significantly, by 4x to 10x, using Nvidia's Blackwell platform combined with optimized software and open-source models for scaling AI applications.
- **Who:** Nvidia, alongside leading inference providers Baseten, DeepInfra, Fireworks AI, and Together AI, reports these dramatic reductions in AI inference expenses.
- **What it does:** It details how combining Nvidia's Blackwell hardware, optimized software stacks, and open-source models dramatically lowers AI inference costs across diverse enterprise production deployments.
- **Why it matters:** This matters because dramatically lower inference costs enable enterprises to scale AI applications from pilot projects to millions of users, making high-volume deployments economically feasible.
- **Strategic view:** TMT executives must recognize that AI inference cost reduction is a multi-faceted strategy combining hardware, software optimization, and open-source models, critical for scalable, economical AI deployment.

☐ Add to Weekly

CEO POV · AI & Space Economy

No CEO statements collected for today.

Patent watch · Compute / Video / Data / Cloud

No relevant patent publications detected for today (EPO / USPTO).

Curated watchlist · 3–5 links per topic

TV / Streaming
No notable articles for this topic today.

Telco / 5G
No notable articles for this topic today.

Media / Platforms
No notable articles for this topic today.

AI / Cloud / Quantum
- MiniMax's new open M2.5 and M2.5 Lightning near state-of-the-art while costing 1/20th of Claude Opus 4.6 (VentureBeat – AI)
  ☐ Add to Weekly
- Nvidia's new technique cuts LLM reasoning costs by 8x without losing accuracy (VentureBeat – AI) ☐ Add to Weekly
- The Download: AI-enhanced cybercrime, and secure AI assistants (MIT Technology Review – Tech) ☐ Add to Weekly
- AI is already making online crimes easier. It could get much worse. (MIT Technology Review – Tech) ☐ Add to Weekly
- What's next for Chinese open-source AI (MIT Technology Review – Tech) ☐ Add to Weekly

Space / Infrastructure
- Highlights of first launch of Ariane 6 with four boosters (ESA – European Space Agency) ☐ Add to Weekly
- Cheops discovers late bloomer from another era (ESA – European Space Agency) ☐ Add to Weekly

Robotics / Automation
No notable articles for this topic today.

Broadcast / Video
No notable articles for this topic today.

Satellite / Satcom
No notable articles for this topic today.