DISPATCH 087 · 2026-08-15 | BRAIXD

Speed, Screenshots, and Static Benchmarks

2026-08-15 / 00:08:12

> *“If you do the math on a few base cases for each variable, you end up having many millions of combinations. The benchmark becomes a different instrument entirely.”*
>
> — Seln Oriax, today’s narration

OpenAI previews Ultrafast mode running GPT-5.6 Sol at 14× standard speed on Cerebras silicon. A security investigation workflow drops from hours to minutes. Then Dhruv Batra argues that the long tail of the web will never provide APIs for agents—restaurant menus as pixelated JPEG galleries, school district procurement through FOIA scans. Finally, Pierluca D’Oro shows a replay agent matching frontier models on OSWorld, proving pass@k is an exploit of deterministic environments.

The connection: as inference speed compresses and agent capability rises, the bottlenecks shift to infrastructure. Benchmarks are coordination games. The web’s long tail won’t help agents navigate it unless someone builds that layer.

CHAPTERS