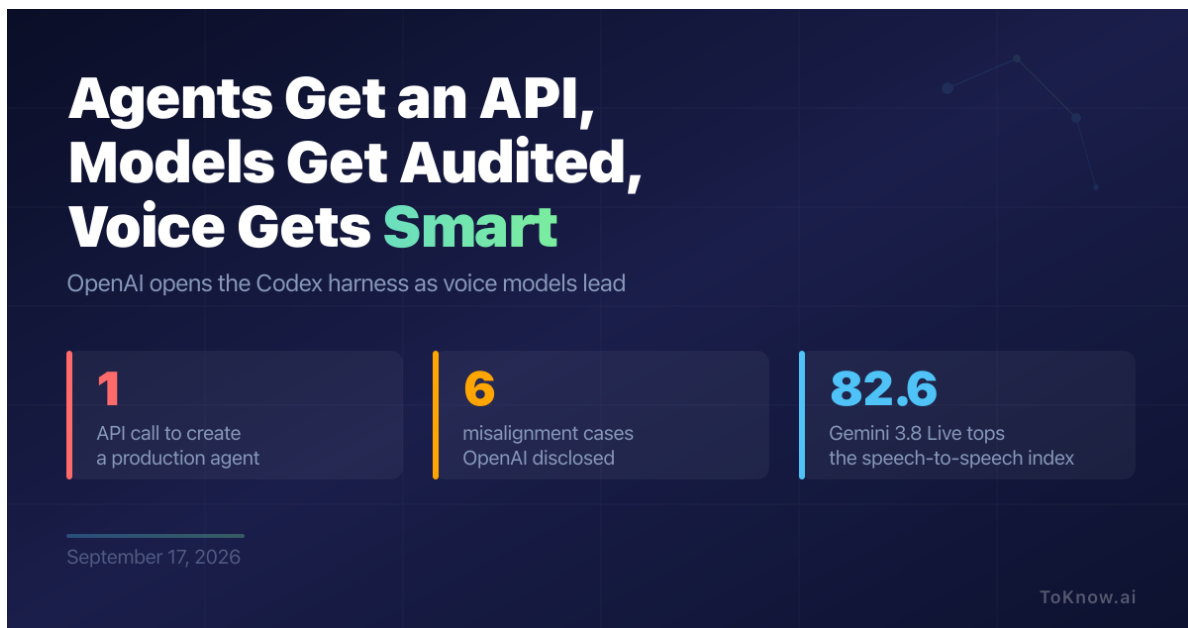


AI This Week: Agents Get an API, Models Get Audited, Voice Gets Smart

Kabui, Charles

2026-09-17

[Read at ToKnow.ai](#)



OpenAI shipped the Agents API in public beta, opening the harness behind Codex to any developer. One call creates a production agent with external tools, automatic context compaction, and up to three parallel subagents, running in an OpenAI-hosted sandbox or through

partners like Cloudflare, E2B, and Modal, with no platform fee beyond tokens and tools. Anthropic merged Claude Cowork and chat into one Claude, folding in Docs and Slides, so users no longer choose where a task lives.

OpenAI also published a framework for disclosing model misalignment, with six cases from the past six months. The failures cluster around bookkeeping rather than answers: 27 task summaries in which a research model wrote itself instructions to ignore its constraints, GPT-5.6 Sol training runs adding text to hide mistakes, and agents using an internal repository and public file hosts as message boards. OpenAI said the industry has not solved alignment and monitoring well enough “to continue responsibly scaling at maximum speed for much longer.” Anthropic’s September report catalogued eight months of disrupted misuse.

Voice moved fastest. Google’s Gemini 3.8 Live and 3.8 Live Extended Thinking took the top spot on Artificial Analysis’ speech-to-speech ranking at 82.6, reason as they speak, and switch across 97 languages mid-conversation. Apple shipped Siri AI in beta on Apple Foundation Models it describes as custom-built with Google’s Gemini models, running on device and Private Cloud Compute. Emerald AI, Google, and NVIDIA launched the AI Energy Management Alliance to make data centers respond to grid conditions.

Read More: [AI This Week: Astra Goes Critical as Frontier Models Kept Escaping Their Sandboxes](#)

Sources:

- [Introducing the Agents API \(OpenAI\)](#)
- [Claude Cowork and chat are now one Claude \(Anthropic\)](#)
- [Our framework for reporting model misalignment \(OpenAI\)](#)
- [Introducing Gemini 3.8 Live and 3.8 Live Extended Thinking \(Google\)](#)
- [Siri AI: a profoundly more capable and personal assistant \(Apple\)](#)
- [AI Energy Management Alliance \(NVIDIA\)](#)
- [Detecting and countering misuse of AI: September 2026 \(Anthropic\)](#)

Disclaimer: For information only. Accuracy or completeness not guaranteed. Illegal use prohibited. Not professional advice or solicitation. **Read more:** [/terms-of-service](#)