Sam Altman Says Astra Will Be the First AI That ’Actually Invents New Things’

2026-08-29T17:20:21.027757Z

Sam Altman Says Astra Will Be the First AI That ’Actually Invents New Things’

body{font-family:-apple-system,BlinkMacSystemFont,’Segoe
UI’,sans-serif;max-width:680px;margin:0 auto;padding:24px;color:#111;line-height:1.7}
h1{font-size:24px;font-weight:700;margin-bottom:8px;line-height:1.3}
.hero{width:100%;border-radius:8px;margin:16px 0 28px;display:block}
p{margin:0 0 18px}
a{color:#0057ff}
.source{font-size:13px;color:#888;margin-top:32px;border-top:1px solid
#eee;padding-top:16px}

Sam Altman Says Astra Will Be the First AI That ’Actually Invents New Things’

In a sweeping profile published by TIME on August 26, Sam Altman previewed OpenAI’s next flagship model, Astra, to a group of VIP customers and made the most consequential claim of his tenure: the model may reach artificial general intelligence.

The quote: "I expect this will be the first model where the model actually invents new things in a way that matters. That’s a very AGI-like thing."

What Altman showed: Astra coordinating multiple agents to jointly complete mathematical proofs, operating desktop software at what Altman described as "superhuman, extremely fast" speeds, generating presentations, financial reports, and structured analysis from messy data. The model is designed to run continuously for days or weeks, remember human corrections, collaborate with other agents and humans, and operate across multiple software tools.

The AGI claim: Altman told TIME that OpenAI may reach AGI - AI that can match or exceed humans at most economically valuable tasks - before the end of 2026. The qualifier: "if you accept his definition." He did not define AGI precisely.

The timing: This preview came in the same week that OpenAI paused Astra development after preliminary evaluations showed it was approaching the ’Critical’ cybersecurity capability tier - the first model in OpenAI’s history to get that close. OpenAI confirmed it was testing Astra under strengthened security controls before resuming work.

What this means: OpenAI is simultaneously publishing a model it considers a potential path to AGI and pausing that model's training because its cyber capabilities are too dangerous to deploy without additional safeguards. Both of those things are true at the same time.

Context for the week: The same seven days produced Anthropic's automated alignment research (weaker AI aligning stronger AI, 15,000? more efficiently than human-led training), OpenAI's Hugging Face breach disclosure, and the Astra Critical tier pause. August 2026 is the most consequential week in AI safety history by several measures - and the frontier models keep shipping anyway.

Source: https://time.com/article/2026/08/26/openai-sam-altman-interview/ (https://time.com/article/2026/08/26/openai-sam-altman-interview/)

https://myspacexai.site/issues/sam-altman-says-astra-will-be-the-first-ai-that-actually-invents-new-things