

# AI Weekly Brief

Saturday, February 8, 2026

Week of Feb 3-8 • 12 major developments tracked

## AI DEVELOPMENT PACE THIS WEEK



🔥 Breakneck — 2 major model drops + Super Bowl AI blitz



## NEW MODEL RELEASES

### Claude Opus 4.6

Anthropic

Anthropic's smartest model yet. Industry-leading across **agentic coding, computer use, tool use, search, and finance** — "often by wide margin." Released Feb 5, 2026. This is the model powering Aiden right now.

Agentic Coding

Computer Use

Tool Use

Finance

Search

### GPT-5.3-Codex

OpenAI

OpenAI's most capable agentic coding model. **First model that helped code itself** — used early versions to debug its own training. 25% faster than GPT-5.2-Codex. New SWE-Bench Pro SOTA. Can build full apps autonomously over days.

Self-Improving

SWE-Bench Pro SOTA

Terminal-Bench SOTA

25% Faster

### Holo2 235B (A22B)

H Company

Takes the lead in **UI Localization** benchmarks. 235B parameter model with 22B active. Focused on visual understanding and UI interaction tasks.

UI Understanding

MoE Architecture

235B / 22B Active

### Nemotron ColEmbed V2

NVIDIA

Top model on **ViDoRe V3** benchmark for multimodal retrieval. Advancing RAG capabilities with state-of-the-art document understanding and embedding.

Multimodal Retrieval

ViDoRe V3 #1

RAG

### SyGra Studio

ServiceNow AI

New synthesis and grading studio for AI model evaluation. Part of ServiceNow's push into enterprise AI tooling.



## PRODUCT &amp; INDUSTRY NEWS



Product

**ChatGPT gets Canva Brand Kit integration** — design directly from ChatGPT using your brand colors/assets. Claude got it first though.

The Verge • Feb 5



Business

**AI.com launches during Super Bowl** — Crypto.com CEO Kris Marszalek pivots to AI, launches personal AI agent platform at ai.com. "Generate a private agent that operates on your behalf."

The Verge • Feb 6



Product

**Gemini becomes AI interior decorator** in Google's Super Bowl ad. "New Home" commercial shows mother and son designing house with Gemini.

The Verge • Feb 6



Product

**Alexa Plus stars in Amazon's Super Bowl ad** — Chris Hemsworth battles AI assistant in comedic spot. AI dominates this year's Super Bowl commercial breaks.

The Verge • Feb 6



Business

**SpaceX / xAI / X public offering could be Musk's cross-company bailout** — xAI burning ~\$1B/month while SpaceX is profitable. Merger talks raise eyebrows about index fund implications.

Bloomberg, WSJ • Feb 4



Tools

**GitHub launches Agentic Workflows** — new framework for building AI agent pipelines directly in GitHub. 169 HN points, major developer discussion.

GitHub • Feb 8



Culture

**Reddit announces bot verification system** — "If you can't distinguish real thoughts from bots, trust erodes." Active work on labeling and authentication.

The Verge • Feb 5



## SAFETY &amp; GOVERNANCE



Safety

**400+ malicious AI skills found on ClawHub** — OpenClaw partners with VirusTotal to scan third-party skills after researchers discover widespread malware in agent add-ons.

The Verge • Feb 6



Safety

**OpenAI poaches safety executive from Anthropic** — Dylan Scardinaro, new "head of preparedness," warns: "The risks of extreme and even irrecoverable harm" require urgent work.

The Verge • Feb 3



Safety

**Sam Altman "kinda-sorta" declares AGI** — "We basically have built AGI, or very close to it." Later walks it back: "I meant that as a spiritual statement, not a literal one."

Forbes via The Verge • Feb 3



## RESEARCH & PAPERS



Paper

**Differential Transformer V2** — Microsoft's follow-up to the viral DiffAttn architecture. Improved attention mechanisms showing strong scaling properties.

Hugging Face Blog • Jan 20



Paper

**"We Got Claude to Build CUDA Kernels and Teach Open Models"** — HuggingFace upskill research: using Claude for RL training to teach open-source models new capabilities. 125 upvotes.

Hugging Face Blog • Jan 28



Paper

**One Year Since the "DeepSeek Moment"** — HuggingFace retrospective on how China's open-source AI ecosystem has evolved architecturally beyond DeepSeek.

Hugging Face Blog • Jan 20–Feb 3



Paper

**Unlocking Agentic RL Training for GPT-OSS** — LinkedIn's practical retrospective on reinforcement learning for open-source GPT models in production agent settings.

Hugging Face / LinkedIn • Jan 27



## CURRENT MODEL LEADERBOARD



### CODING

- |    |                 |     |
|----|-----------------|-----|
| 1. | GPT-5.3-Codex   | NEW |
| 2. | Claude Opus 4.6 | NEW |
| 3. | GPT-5.2-Codex   | —   |
| 4. | Sonnet 4.5      | —   |



### REASONING

- |    |                  |     |
|----|------------------|-----|
| 1. | Claude Opus 4.6  | NEW |
| 2. | GPT-5.3-Codex    | NEW |
| 3. | Gemini 2.0 Ultra | —   |
| 4. | DeepSeek V3      | —   |



### VISION / UI

- |    |                 |     |
|----|-----------------|-----|
| 1. | Holo2 235B      | NEW |
| 2. | Claude Opus 4.6 | NEW |
| 3. | GPT-5.2         | —   |
| 4. | Gemini 2.0 Pro  | —   |



### RETRIEVAL / RAG

- |    |                      |     |
|----|----------------------|-----|
| 1. | Nemotron ColEmbed V2 | NEW |
| 2. | Cohere Embed V4      | —   |
| 3. | OpenAI text-embed-3  | —   |
| 4. | BGE-M3               | —   |



## OV'S EDGE — WHAT THIS MEANS FOR YOU

**Massive week.** Two frontier model drops (Opus 4.6 + GPT-5.3-Codex) in the same week means the coding agent war is heating up fast. You're already running both — that's an advantage over 99% of people.

**GPT-5.3-Codex helping code itself** is the headline. Self-improving AI models are no longer theoretical — OpenAI deployed it in production. This accelerates everything.

**Super Bowl was an AI showcase** — Google, Amazon, and AI.com all ran AI ads. This means mainstream adoption is accelerating. Your clients (like AdTime) need to hear this: AI isn't

coming, it's here. Use this in your next pitch.

**ClawHub malware warning** — 400+ malicious skills found. Be careful with third-party OpenClaw skills. Stick to verified sources.

---

🔑 Already running Opus 4.6 + Codex 5.3 — you're on the frontier

📺 Use Super Bowl AI momentum in AdTime pitch — "AI went mainstream this weekend"

👁️ Watch GitHub Agentic Workflows — could replace some of our custom agent tooling

⚠️ Audit any third-party OpenClaw skills you've installed

📖 Read the GPT-5.3-Codex self-improvement blog post — it's a paradigm shift