

Microsoft announces seven MAI models trained in-house, including MAI-Thinking-1

Source copy kept by wheresthe.ai for the story "Microsoft announces seven MAI models trained in-house, including MAI-Thinking-1" (<https://wheresthe.ai/d/microsoft-seven-mai-models-2026-06/>).

Source: <https://microsoft.ai/news/building-a-hillclimbing-machine-launching-seven-new-mai-models/>

Text saved 18 Oct 2026 (the page as published).

SHA-256 of the saved text: de81a88e0eef693e55f575d258d13303024e7498751d8c8d2dbd3e6a7393c6c1

Copilot -->

Announcements

Building a hill-climbing machine: Launching seven new MAI models

Mustafa Suleyman

June 2, 2026

Announcements

Mustafa Suleyman

LI

X

FB

Updated as of June 8, 2026.

Today we are announcing a family of seven new models developed in-house at Microsoft AI. Beyond these models, we're building a superintelligence lab – a system and an approach we believe will define the next phase of AI.

This is an extraordinary time in technology. The compute used to train frontier models has increased by a factor of one trillion . Now we expect another thousand-fold increase over the next three years, which in turn means more advanced capabilities, and the continued rollout of ever more effective AI.

This epic compute ramp will change the nature of work, business and daily life. We all have to prepare for this reality. Our job at MAI is to help you do this – to push the frontier, and to build a hill-climbing machine to keep you at the frontier.

Here's our first steps along the way.

Our Models

Our new models across image, voice, transcription, coding, and reasoning, together form the MAI model family: a multimodal ecosystem designed to work across the kinds of tasks that matter in the real world.

MAI-Thinking-1 , Microsoft AI's flagship reasoning model. It is a medium-sized model that stands among the strongest models in its weight class: it matches leading models on key software engineering benchmarks, and demonstrates advanced mathematical reasoning capabilities, and is preferred to Sonnet 4.6 in our blind human side-by-side evaluations. We trained it from the ground up on clean data, without distillation from third-party models.

MAI-Code-1-Flash is an inference-efficient agentic coding model. This model is tailor-made for and deeply integrated into GitHub Copilot, VS Code and the Microsoft stack, and, with 5 billion active parameters, is comparable to Haiku but cheaper.

MAI-Image-2.5 including its ultra-efficient Flash variant, supports both world-class text-to-image and image editing, surpassing the Arena score of Nano Banana Pro.

MAI Transcribe-1.5 is the best transcription model in the world, with SOTA accuracy. It's five times faster than competing models, with built-in support for domain-specific terminology across 43 languages.

MAI-Voice-2 brings high-quality, natural-sounding speech generation across 15 languages, with the ability to adapt to a voice from a short sample, alongside strong safeguards against misuse. MAI-Voice-2-Flash, coming soon, does it in a lower cost, ultra-efficient package.

Alongside distribution on Foundry and optimization for our 1P products, our models are also going to be widely available for developers on OpenRouter, as well as Fireworks and Baseten. For the first time developers will be able to tune the weights of the model themselves.

All these models share the same infrastructure and the same commitment to clean, enterprise grade data lineage. We don't distill from other labs and we don't rely on opaque data. Our datasets are clean, traceable, and enterprise-grade. They are designed to work together, and to integrate directly into the products people use every day. But the models themselves are only part of the story.

The most important shift lies in what you can do with them.

Adapted for you

AI is moving into a new phase. With reinforcement learning in real-world environments, AI can fully adapt to the specifics of a given workflow for the first time. We call this Microsoft Frontier Tuning. We think it's the future for how AI shows up. Learn more in the Microsoft 365 blog post.

We think it's the future for how AI shows up.

In this set-up the most valuable data is yours: the trace of real work an agent completes, the sequence of steps, the decisions, the actions taken that define how tasks actually get done inside an organization.

Our reinforcement learning environments (RLEs) allow your MAI models to learn directly from your workflows. Think of them as training gyms for AI, accessible only to you.

With Frontier Tuning, you're building your own model, trained on your data, within your environment, controlled by you. Your institutional knowledge becomes part of the model, and it stays yours. What's even better is that this adaptation then drives efficiency and performance.

Across Microsoft and with customers, Frontier Tuning is showing that custom models are both better and more efficient: our MAI tuned model for Excel matches GPT 5.4 while being up to 10x more efficient. Early adopters are seeing similar gains at the frontier. When tuned for a market-leading organization's exacting enterprise standards, MAI achieved the highest win rate of any model tested at roughly 10x lower cost.

Developers and businesses have been crying out for AI that delivers on their terms and under their say. We see this as a major step towards delivering that.

Frontier health intelligence with the Mayo Clinic

There are a number of high importance, high sensitivity domains – like health – that require even more intense collaboration. That's why today we're also announcing that Microsoft and Mayo Clinic are collaborating to co-create a frontier AI model for healthcare that brings together Mayo Clinic's world-leading clinical expertise, de-identified clinical data and longitudinal insights with Microsoft's foundational AI capabilities.

This model will be designed to excel at the broadest scope of clinical reasoning and healthcare use cases, reaching a level that today's general-purpose systems simply cannot match.

The model will first be deployed within Mayo Clinic's own environment, the world's top hospital system, where we expect it to enable a broad range of capabilities, including earlier and more accurate diagnoses and treatment planning. Once validated, the model will be made available to other organizations via Microsoft Foundry, making Mayo Clinic's expertise accessible to many more who need it.

The frontier AI model will be owned by Mayo Clinic, reinforcing our mutual longstanding commitment to patient trust, clinical rigor, safety and responsible stewardship of clinical health data and AI.

Our Lab

At Microsoft AI, we recognize that there are no shortcuts to the frontier. We train our reasoning models from

scratch. We don't distill from other labs and we don't rely on unlicensed or opaque data. Our datasets are clean and appropriately licensed. Every component of the system, from architecture to training pipeline to post-training, we built ourselves. We co-design with our own Maia 200 silicon, and are already seeing a 1.4x efficiency boost from these efforts. This is all about long term self-sufficiency for Microsoft and our partners. It's about models you can trust.

The goal here is to build what we think of as a hill-climbing machine: an organization that can continuously improve, cycle after cycle, as we apply more compute, better data, and sharper evaluation.

We think scientific rigor is critical to this. That's why for everything we do, we ablate, we measure, we document. We invest heavily in data pipelines. We work in small teams with falsifiable goals over short time periods, matching velocity with quality, focus with ambition. And we are committed to transparency. We want to bring you with us on this journey. That's why today we are publishing in depth safety and technical reports.

Humanist Superintelligence

This is what we are building in MAI: a family of models, with new versions available now. A lab built on first principles, focused on long-term capability. And a new approach to tuning and ownership that we believe will define the next phase of AI.

Our ultimate goal is what we call Humanist Superintelligence. That means advanced AI systems designed to serve people and organizations, not replace them. These systems must remain tools, shaped by human intent, accountable to human oversight, and ultimately subordinate to human goals. People – you – must always remain in control.

Over the next year, stand by for a rapid scale-up in our compute and capabilities as we push to make this ambition a reality. It's a new phase for AI, and for us.

– The MAI Team

Build the Future With Us

We're a lean, fast-moving lab made up of some of the world's most talented minds. We have an exciting roadmap of compute at MAI, with our next-generation GB200 cluster now operational. And we have an ambitious mission we truly believe in. We're also fortunate to partner with incredible product teams giving our models the chance to reach billions of users and create immense positive impact. If you're a brilliant, highly-ambitious and low ego individual, you'll fit right in—come and join us as we work on our next generation of models!

[Explore all jobs](#)

[Related Stories](#)

[Microsoft Build 2026: MAI keynote transcript](#)

[build](#)

[Our values in operation: Health](#)

[partnerships](#)

[Introducing MAI-Thinking-1](#)

[models](#)