

Inkling adds another name to open-weight AI

Friday, July 17, 2026

Today's top story, from reporter Sascha Brodsky:

AI startup Thinking Machines expanded the growing ranks of open-weight AI models Wednesday with Inkling, the first release in a new model family.

Founded by former OpenAI CTO Mira Murati, the company pitched Inkling as a model developers could shape for their own needs, rather than one designed to lead every benchmark. It's one of a growing number of open model families, including IBM Granite, looking to give developers more control over how they run and adapt AI systems.

"I've played with it. It's pretty good," Chris Hay, a Distinguished Engineer at IBM, said on this week's Mixture of Experts podcast. "They've made interesting architectural choices. It's not the same model as every other model."

Inkling contains 975 billion parameters, the numerical settings an AI model learns during training, but activates only 41 billion at a time. Its mixture-of-experts design directs each input to the parts of the model best suited to handle it, helping reduce the computational power required to produce a response.

Thinking Machines released Inkling under an Apache 2.0 license, allowing developers to download, run and fine-tune the model. Open-weight systems make their trained parameters available, unlike closed models, which keep those inner workings private and usually require customers to use a company's hosted service.

The startup also paired Inkling with Tinker, its fine-tuning platform. The combination allows developers to retrain the model for specific tasks, rather than rely solely on a general-purpose system.

Thinking Machines acknowledged that Inkling was "not the strongest" model available. Instead, the company emphasized its ability to handle text, images and audio, along with its support for customization.

That greater control could prove more useful to some businesses than winning a leaderboard, IBM Fellow Aaron Baughman said on the podcast.

"The future of AI might not be about building the biggest model," Baughman said, "but about giving people the ability to customize, adapt and control these types of models for their own needs."