

Open Source This Month: Video With Sound, a Trillion-Parameter Model, and a Moderator That Reads Your Rules

Kabui, Charles

2026-10-13

 Read at [ToKnow.ai](#)



The banner features a dark blue background with a network diagram of three nodes in the top right. The main title is in large white and green font. Below it is a subtitle. Three vertical bars (orange, yellow, and blue) separate three key statistics. The date and ToKnow.ai logo are at the bottom.

**Open Source This Month:
Video With Sound and a
Trillion-Parameter Model**

From trillion-parameter media models to phone-sized specialists

1.02T parameters in MiMo-V2.6 Pro, 42B active per token	44 kHz synchronized audio from Kandinsky 6.0 Video	740M parameters in EmbeddingGemma 2, runs on a phone
--	---	---

October 13, 2026 ToKnow.ai

October's open-source releases split between large models that handle many kinds of media and small specialists. Xiaomi's [MiMo-V2.6](#) is a family that handles text, audio and video,

led by a 1.02-trillion-parameter model that activates 42 billion per token, alongside a 310-billion-parameter Flash version, and Xiaomi released the reinforcement-learning environments it trained on, spanning code, cyber, general, visual and music tasks. [Kandinsky 6.0 Video](#) generates five-second clips with synchronized 44 kHz audio, lip-sync included, from 3-billion and 29-billion-parameter models that upscale to Full HD. [YuE2](#) writes a readable score before rendering it as a full song, and expert listeners preferred that planning step by 49.3% to 34.6%.

The smaller releases show where open models are heading: narrow, cheap jobs. Google's [EmbeddingGemma 2](#) is a 740-million-parameter model that turns text, images, audio and video into one common set of numbers, so a single search can match across all four, and it runs on a phone in about 567 megabytes. [PolicyLM-1.7B](#) is a moderation classifier that reads a platform's own rules with every message, returns a score from 0 to 1 for each category, and answers in under 100 milliseconds, with no retraining when the rules change.

Open models are dividing the work. The largest compete on scale, and increasingly ship the training recipe, not just the weights. Small models take the narrow jobs, small enough to run on a phone or a single graphics card.

Read More: [the previous open-source radar, on an open code corpus and audits that check themselves](#)

Sources:

- [MiMo-V2.6: Scaling Reinforcement Learning Towards Self-Improvement \(arXiv\)](#)
- [Kandinsky 6.0 Video: Foundation Models for Synchronized Video and Audio Generation \(arXiv\)](#)
- [YuE2: Unifying Symbolic and Audio Music Generation at Frontier Quality \(arXiv\)](#)
- [EmbeddingGemma 2: an open, lightweight multimodal embedding model \(Google\)](#)
- [EmbeddingGemma 2 model card \(Hugging Face\)](#)
- [Introducing PolicyLM-1.7B \(Musubi\)](#)

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