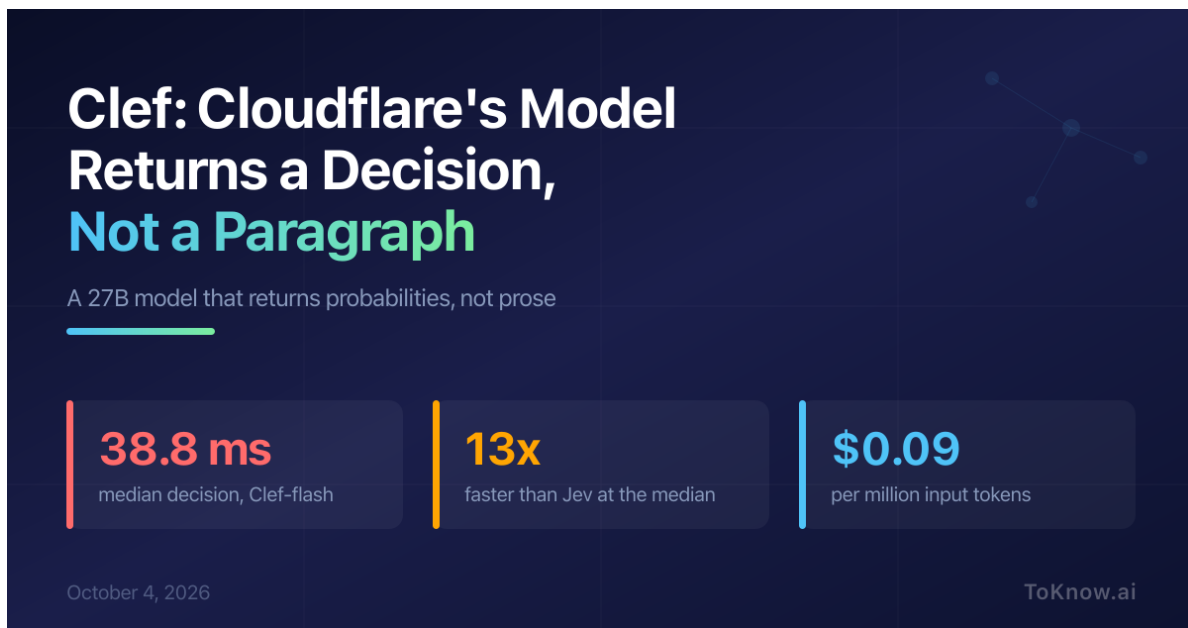


Clef: Cloudflare's Model Returns a Decision, Not a Paragraph

Kabui, Charles

2026-10-04

[Read at ToKnow.ai](#)



Clef is an AI model that answers with a probability instead of a sentence. Cloudflare released two versions, the 27-billion-parameter Clef and the faster 9-billion-parameter Clef-flash, both post-trained from [Alibaba's Qwen family](#). It is built for software that has to decide: is this support ticket urgent, which team should handle it, is this domain a phishing site. You send text, JSON, an image, or a video, plus a short list of typed questions and the allowed answers.

Clef returns a probability for every option in a single pass, with no paragraphs to parse and no reasoning to wait for. Both are downloadable, run on Cloudflare's Workers AI, and speak the same API as TypeSafe's Jev, so teams can swap them in.

Speed is the reason it exists. Clef-flash answers in a median 38.8 milliseconds, about 13 times faster than Jev's 524 milliseconds, fast enough to sit inside an agent's decision step rather than behind it. You pay only for input tokens, \$0.24 per million for Clef and \$0.09 for Clef-flash, because there is no generated text to bill. Cloudflare's threat team pointed it at website domains, where a full fetch, render, and classify cycle took 2.2 seconds against 4.7 seconds for the general model it compared against.

Six weeks ago the decision model was a startup's pitch. Now a major cloud provider ships an open competitor, top on seven of the ten decision benchmarks it measured, with a fine-tuning service for training it on your own data.

Read More: [Jev: The AI Model That Answers in Decisions, Not Sentences](#)

Sources:

- [Introducing Clef: our open-source decision models, and new RL fine-tuning platform \(Cloudflare\)](#)
- [Cloudflare/clef model card \(Hugging Face\)](#)
- [clef: model specs and pricing \(Cloudflare Workers AI docs\)](#)
- [Introducing Clef: Cloudflare's first open-source decision models, now on Workers AI \(Cloudflare Changelog\)](#)
- [Cloudflare Releases Clef and Clef-flash: Open-Weight Decision Models That Return Typed Probabilities Instead of Text \(MarkTechPost\)](#)

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