

Mythos Debate Sharpens as Meta Launches Muse Spark and Open Models Advance

AI News Digest

2026-04-09

Mythos Debate Sharpens as Meta Launches Muse Spark and Open Models Advance

By AI News Digest • April 9, 2026

Debate over Anthropic’s Mythos shifted from alarm to questions of evidence, diffusion, and governance. Meta launched Muse Spark, while new releases and adoption data pointed to faster movement in the open-model ecosystem.

The Mythos debate moved from alarm to evidence

Cyber risk looks real, but the size of the step is contested

Anthropic’s unreleased Mythos is being described by briefed officials and commentators as a potentially dangerous cyber model, and Gary Marcus argued the episode strengthens the case for government oversight rather than leaving release decisions to company leaders [1, 2, 3]. But the claims are already being challenged: Heidi Khlaaf flagged missing comparison benchmarks as a red flag, and Marcus said Mythos may not be as bad as the reporting suggests even if it still could cause harm without needing to qualify as AGI [4, 3].

Why it matters: The conversation is moving away from “is this AGI?” and toward a more practical question: how cyber-capable models should be evaluated, released, and governed [3].

Open models already reproduce parts of the showcase

A follow-on analysis shared by Clement Delangue found that eight out of eight small, cheap open-weight models detected Mythos’s flagship FreeBSD exploit, including a 3.6B-active model costing \$0.11 per million tokens; a 5.1B-active open model also recovered the core chain of a 27-year-old OpenBSD bug [5]. Another post summarized the broader result as a “super jagged” frontier, with rankings reshuffling across tasks rather than one model dominating everything

[6]. Martin Casado said models getting better at vulnerability finding could be positive if it lowers the cost of discovery and reduces zero-day hoarding [7, 8].

“The models are ready. The question is whether the rest of the ecosystem is.” [9]

Why it matters: If useful cyber capability is already diffusing into smaller open models, defenders may need to focus less on a single frontier release and more on integrating these tools into real workflows now [9].

Meta turned its rebuilt AI stack into a product

Muse Spark is now live in Meta AI

Meta introduced Muse Spark, the first model from Meta Superintelligence Labs, describing it as a natively multimodal reasoning model with tool use, visual chain of thought, and multi-agent orchestration [10]. It is available today in Meta AI and the Meta AI app, with a private-preview API for select partners, and Meta said future versions may be open-sourced [10]. Meta also said the model shows competitive performance in multimodal perception, reasoning, health, and agentic tasks, while it continues investing in long-horizon agents and coding workflows where it sees current gaps [11].

Why it matters: This is Meta’s first public model from its new superintelligence lab, and the company is shipping it as a product while positioning larger models as the next step [10, 11].

Meta’s bigger claim is about efficient scaling — and that is already being debated

In a technical thread, Meta said a rebuilt pretraining stack can reach the same capabilities with over an order of magnitude less compute than Llama 4 Maverick, and that its RL stack delivers smooth gains plus more token-efficient reasoning through thinking-time penalties and multi-agent orchestration at comparable latency [12, 13, 14, 15]. Meta is also rolling out Contemplating mode, which it says uses parallel agents to compete with the extreme reasoning modes of Gemini Deep Think and GPT Pro [16]. François Chollet pushed back, arguing the new model already looks overoptimized for public benchmarks at the expense of actual usefulness [17].

Why it matters: Meta is not just launching a model; it is making a broader claim that its new stack scales efficiently. The immediate pushback shows how central the benchmark-versus-utility debate has become [12, 17].

Open-weight competition keeps shifting toward coding agents and Chinese adoption

GLM-5.1 makes a strong bid for the top open-weight coding model

Z.ai launched GLM-5.1 and said it ranks #1 among open-source models and #3 globally on SWE-Bench Pro, Terminal-Bench, and NL2Repo [18]. The company said the model is built for long-horizon tasks, with autonomous runs up to eight hours and thousands of refinement iterations, while Sebastian Raschka described it as a DeepSeek-V3.2-like architecture with more layers and called it “THE flagship open-weight model now” based on the published benchmarks [18, 19].

Why it matters: The open-weight race is getting more focused on sustained coding and agent execution, not just chat quality [18, 19].

New adoption data points to continued momentum for Chinese open models

The new ATOM Report says Chinese models are continuing to accelerate in open-model adoption and hold a strong lead in derivative models and Open-Router inference share [20, 21]. Its RAM metric highlighted Qwen 3.5, Nemontron 3, and Kimi K2.5 as standout recent models, based on a manually curated set of roughly 1,500 important language models [20, 21]. Google, meanwhile, said Gemma 4 passed 10 million downloads within a week of launch, taking the Gemma family past 500 million total downloads [22].

Why it matters: Distribution remains broad, but the newest adoption data suggests Chinese model families are still gaining share quickly inside the open ecosystem [20].

Two quieter infrastructure signals worth watching

Anthropic productized long-running agents

Anthropic announced Managed Agents, a hosted service for long-running agents, and framed the engineering challenge as designing systems for “programs as yet unthought of” [23].

Why it matters: Labs are increasingly packaging agent runtime infrastructure as a product, not just releasing stronger base models [23].

Safetensors moved deeper into the core AI stack

Hugging Face said Safetensors, created with collaborators including EleutherAI and Stability AI, has become the most popular way to share models safely and is now joining the PyTorch Foundation, with further scale-up including possible torch-core integration [24].

Why it matters: Secure model distribution is becoming part of the default ecosystem plumbing, not a side project [24].

Sources

1. X post by @JimVandeHei
2. X post by @GaryMarcus
3. X post by @GaryMarcus
4. X post by @HeidyKhlaaf
5. X post by @ClementDelangue
6. X post by @stanislavfort
7. X post by @martin_casado
8. X post by @martin_casado
9. X post by @ClementDelangue
10. X post by @AIatMeta
11. X post by @AIatMeta
12. X post by @AIatMeta
13. X post by @AIatMeta
14. X post by @AIatMeta
15. X post by @AIatMeta
16. X post by @AIatMeta
17. X post by @fchollet
18. X post by @Zai_org
19. X post by @rasbt
20. X post by @natolambert
21. X post by @natolambert
22. X post by @sundarpichai
23. X post by @AnthropicAI
24. X post by @ClementDelangue