The fear around Mythos is also self-serving AI hype: For India, safety might lie in more open source resources

Indian Express | 2026-04-29

Pranesh Prakash

Relying on proprietary LLMs for national security is an unacceptable supply-chain risk.

"India seeks fair access to Anthropic's Mythos for critical infrastructure security" was a headline yesterday. The 'Mythos' being referred to is Anthropic's upcoming proprietary large language model (LLM), which they have deemed too dangerous to release, because it apparently excels at discovering security vulnerabilities in software. Under 'Project Glasswing', Anthropic has provided exclusive access to the US government and a select set of US software companies and projects, in order to uncover and patch vulnerabilities. The reaction in New Delhi has been swift, but deeply misguided. Some tech policy wonks have urged India to "urgently seek participation in Project Glasswing." Pleading with the US for access to proprietary American tech concedes digital sovereignty, and won't help our cyber-defences. To understand why, we must deconstruct the Mythos hype.

It is true that frontier LLMs completely upend traditional cybersecurity—but not in the way Anthropic's hype suggests. While Mythos has been able to find security problems in important software projects like Linux, OpenBSD and Firefox, doing so took thousands of attempts, aside from a lot of money—Anthropic has set aside USD 100 million for the project. A report by Aisle, a cybersecurity firm, found that they could independently replicate many of the publicized findings by Mythos using smaller, cheaper, open-weights models. A single Linux developer using an open-source LLM on a laptop has found several dozen vulnerabilities so far with a custom tool. A paper by Liu et al., showed a coordinated group of open source models (Kimi K2.5, in this case) could find ten previously unknown ("zero-day") vulnerabilities in Google Chrome. It demonstrated that while Anthropic's latest proprietary model was highly capable, deploying open-source models locally was the only economically viable way to bypass the prohibitive costs and limits associated with massive-scale, continuous vulnerability scanning.

Importantly, finding zero-day vulnerabilities is fundamentally different from weaponizing them at scale, especially if good systems engineering principles like

defence-in-depth are practiced. Many of the vulnerabilities found by Mythos weren't exploitable. In the speed-race of modern cybersecurity—detection, verification, coordination, and patching across systems—access to powerful open-source LLMs that can be modified and run cost-efficiently at scale is far more critical than having the absolute best model at any moment. Most successful cyber-attacks are from known-but-unpatched vulnerabilities along with lack of defence in depth, rather than the zero-day vulnerabilities that are at the core of the Mythos hype. (Much to Anthropic's chagrin, Mythos couldn't prevent unauthorized people from gaining access to Mythos!)

Will those seeking to exploit vulnerabilities (the US's NSA, India's NTRO, criminals, etc.) gain the upper hand because of LLMs or those seeking to defend against them? Clement Delangue, Huggingface's CEO, points out that 'the Mythos moment' is asymmetrical in its impact: it helps the defenders more when it comes to FOSS, but attackers more when it comes to proprietary software. Historically, proprietary software relied on 'security through obscurity': hoping that obscuring the source code makes software harder to exploit. FOSS (and all of cryptography) relies on openness for security, captured in the dictum "given enough eyeballs, all bugs are shallow". Delangue argues that since LLMs can now read stripped binaries (the part that proprietary software cannot hide), legacy proprietary firmware running in critical information infrastructure (CII) is suddenly legible to automated analysis and attacks. But FOSS is better protected since independent developers can use diverse AI toolchains (as long as they aren't regulated away) to investigate and fix bugs—"given enough eyeballs" simply evolves into "given enough eyeballs and AI agents and computing power." Just as previous generations of bug-finding automation, like fuzzing, were adopted by software developers and security professionals, they will now need to use LLMs as well; not doing so will give a leg up to the attackers. The security hype around LLMs is largely true; the hype around Mythos isn't.

US Big Tech companies and officials have compared advanced AI models to digital nukes, hoping to erect regulatory barriers to competition from Chinese and Indian companies—who are mostly working with and releasing open-weights (and in many cases, fully open source) AI models. Buying into that would harm our cybersecurity posture by lowering access to open AI models, and slowing down innovation and the spread of the benefits of AI.

If India's CII depends on proprietary American LLMs like Mythos, our cyber-immune system can be revoked by a whim of US foreign policy. 'But isn't China a potential threat as well¿ one might ask. Yes, it is, but the reality is that this is not a conflict between US and Chinese tech; it is a battle between dependency versus the digital sovereignty that FOSS and open-source AI systems enable—once available under an open licence, code and models can be modified to suit one's own needs and run locally. Relying on proprietary LLMs (optimized for a single country's chip architecture) for our national security is an unacceptable supply-chain risk. The idea that only access to proprietary frontier models can

secure our infrastructure is not only demonstrably false, but is the inverse of the truth.

In 2019, OpenAI initially held that GPT-2 was too dangerous to release, though that was clearly false. Now Anthropic is saying the same thing. We should learn to ignore such self-serving hype, and urgently push for FOSS and open models for the sake of our digital sovereignty and security.

Pranesh Prakash is a tech law and policy analyst, and consults with think tanks, tech companies and universities. The views expressed are personal, and may conflict with those of his clients.