Quote by Devroop Dhar, Co-founder and India CEO, Primus Partners

Published in Analytics India Magazine
May 04, 2026 | 7:00 PM

Anthropic Mythos Isn't Out Yet, But It's Already Exposing India's AI Governance Gaps

India's current legal architecture is insufficient to address AI-powered cybersecurity threats.

Authored by Shalini Mondal

Read on: Anthropic Mythos Isn't Out Yet, But It's Already Exposing India's AI Governance Gaps | Devroop Dhar, Primus Partners, AI cyber threats, banking cybersecurity, AI phishing attacks, identity verification, behavioural analytics, transaction risk assessment

Article Content:

Anthropic hasn't even released Claude Mythos to the general public, but it's already spurring governments to action.

In a recent meeting with bank leaders and financial services stakeholders, Finance Minister Nirmala Sitharaman flagged emerging risks from advanced AI models such as Claude Mythos, warning that India's banking system may need "something far more versatile" to counter evolving cyber threats.

Anthropic's Claude Mythos is designed to enhance AI-driven cybersecurity by detecting system vulnerabilities. However, its advanced capability can also identify weaknesses and suggest potential exploitations. This has sparked global concern, with experts warning that such tools could be misused to target and disrupt critical systems, including

banking infrastructure, as well as critical digital infrastructure such as electricity grids, water systems, and public governance platforms.

For India's policymakers, the issue is urgent. As advanced AI systems become increasingly capable of autonomous decision-making, India faces a narrowing window to build the legal and cybersecurity frameworks needed to protect its governance infrastructure.

"A program like Mythos can prejudicially impact governments, governance, and also their sovereignty, both in the physical world and also cyber sovereignty and AI sovereignty," Pavan Duggal, a Supreme Court advocate specialised in cyberlaw, cybercrime, and cybersecurity, tells AIM.

The concern stems from the ability of such systems to autonomously identify and exploit cybersecurity vulnerabilities, including legacy flaws that may have remained undetected for decades.

"Mythos, as an algorithm, is empowered enough to find cybersecurity vulnerabilities in existing systems, including discovering vulnerabilities which may be more than 20 years old," he comments.

Inadequate Cyber Resilience

Even US officials refused a wider rollout of Mythos as this could heighten security risks. Anthropic had proposed increasing access to Mythos from roughly 50 organisations to around 120, including companies tied to critical infrastructure, according to a Wall Street Journal report.

Meanwhile, developing economies like India are particularly vulnerable due to weaker cyber resilience frameworks and underdeveloped protection mechanisms for critical information infrastructure. Despite the threats, the Centre is working on ways to enable Indian companies to access Mythos, according to an Economic Times report.

According to the 2026 Cyber Security Report published by Check Point Software, Indian organisations faced an average of 3,195 cyber incidents per week in 2025—a 2% increase from the previous year. Of them, government bodies are among the most vulnerable, with nearly 5,000 weekly attacks per week.

Duggal notes that while governments have traditionally focused on threats from state and non-state actors, autonomous AI systems introduce a fundamentally different risk.

Meanwhile, when it comes to banking, Devroop Dhar, Co-founder and India CEO of Primus Partners, clarifies that Claude Mythos-type programmes don't actually break vulnerable bank systems. “While basic banking systems, encryption algorithms, and procedures for secure transactions are not affected, the new challenge is in the manner of conducting attacks against them,” he notes.

AI brings a new level of sophistication to cyberattacks since adversaries can create complex contextual and timely interaction models of banking workflows, from phishing to support calls that would encourage users to take some actions, thereby removing barriers that existed before.

Another issue is automation. Instead of engaging one-on-one with the target, hackers can launch mass attacks with automated content that adapts to responses. Traditional methods of pattern recognition will become obsolete.

The problem is not with banking itself, but with the trust layer. “Banks have to understand that perimeter security won't help them any more in terms of security. They need to focus on verifying identity, analysing users' behaviour, and transactional risk assessment. Properly secured authentication measures will play a key role,” Dhar explains.

AI Governance Gap Widens in India

Duggal argues that India’s current legal architecture is insufficient to address AI-powered cybersecurity threats and flags the need for a national cybersecurity strategy.

“We have only had the Information Technology Act 2000... and we only added very cosmetic provisions on cyber security under the IT Amendment Rules 2025,” he explains.

India’s National Cyber Security Policy 2013, he says, has failed to translate into actionable implementation. The policy led to the creation of a 24x7 National Critical Information Infrastructure Protection Centre to protect critical infrastructure, as well as strengthened CERT-IN for crisis management.

Beyond cybersecurity, the absence of a dedicated AI law remains a major concern.

India’s ‘seven sutras for AI governance’, released by the Ministry of Electronics and IT in November 2025, are seen as insufficient. “Those are broad, generic principles,” he states. “What India now requires is a policy which should have the strength of the law.”

Without legal enforceability, voluntary frameworks lack deterrence. “There are no legal penal consequences should you not go ahead and follow any of the seven sutras,” Duggal adds.

The rapid evolution of generative and agentic AI has only intensified the debate to regulate the technology. As AI agents begin making autonomous decisions, legal systems must determine responsibility when harm occurs. Once AIbegins causing harm to humans, the fundamental issue that comes up for consideration is who should be made accountable.

Duggal stresses that India needs principle-based legislation capable of adapting to technological shifts, rather than rigid laws requiring constant amendment. “You cannot have one AI law and keep on amending it every month because AI is moving at a rapid pace,” he states.

Instead, he advocates for broad legal principles backed by secondary legislation powers.

Addressing concerns that stricter regulation could stifle innovation, he pushes back strongly. “There’s often a misnomer in the minds of policymakers that if we regulate, that’s going to impede innovation,” Duggal explains. He argues that clear legal frameworks are essential for protecting innovators and encouraging long-term technological growth.

Rather than copying global frameworks such as the EU AI Act or China’s stateled governance approach, India should craft a model tailored to its own realities. His strongest warning was directed at policymakers delaying action.

He warns that once AI systems surpass human oversight capabilities, meaningful regulation may become significantly harder. Duggal calls on the government to proactively safeguard what he described as citizens’ cognitive rights.

“The time has come for the government to stand up and say: AI should not be in a position to pollute the mental mindsets and also the cognitive faculties of our citizens,” he states.

As AI systems continue to evolve at breakneck speed, the question for India is no longer whether regulation is necessary, but whether it can move fast enough to shape the rules before the technology begins shaping them.