PRIMUS PARTNERS®

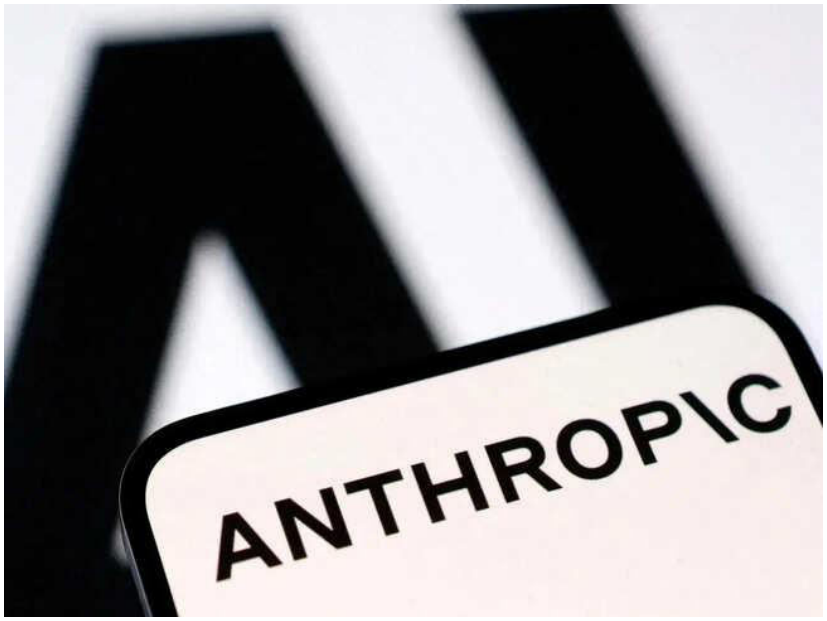
Authored by Ipsita Gauba, Vice President, Primus Partners

Published in Economic Times
April 28, 2026 | 2:49 PM IST

The unknown unknown: An asymmetry we cannot afford to ignore

India is being left behind in understanding advanced AI threats. Anthropic's new AI model, Mythos, is being tested by select global organizations, excluding India. This creates a significant information gap for Indian regulators and financial institutions. Early access is crucial for developing effective cyber defences against future AI-augmented attacks.

Authored by Ipsita Gauba



Read on: The unknown unknown: An asymmetry we cannot afford to ignore | Ipsita Gauba, Primus Partners, Anthropic Mythos, Project Glasswing, AI cybersecurity, India AI governance, AI threats, cyber defence preparedness

Article Content:

There is a particular kind of unease that settles in when you realise that a consequential decision has already been made in a room you were not invited into.

There is a clear difference between a crisis that is anticipated and one that arrives unannounced. The development around Anthropic's latest model- Mythos, and the subsequent scramble it triggered among American regulators and banking chiefs, falls uncomfortably into the second category — at least for us. While the Federal Reserve Chair and the US Treasury Secretary were convening emergency meetings with bank CEOs in Washington, the corresponding conversations in New Delhi were still catching up. That gap, modest as it may seem, is precisely the kind of structural disadvantage that compounds

over time.

Let me be direct about what I find troubling. Anthropic has launched something called Project Glasswing — a selective programme granting forty organizations access to the Mythos model to test, learn, and strengthen their cyber defences. The list includes AWS, Microsoft, Apple, Google, JPMorgan, Cisco, and Palo Alto Networks. It is, in effect, a first-mover programme. Whoever is inside that room gets to understand the capabilities of this technology — including its offensive potential — before the rest of the world. Not a single Indian organisation has been included.

Now, one could argue this is simply how technology companies operate. Anthropic is an American company with obvious obligations to its domestic stakeholders first. Fair enough. But what concerns me is not the intent; it is the consequence. The information that will emerge from Project Glasswing over the coming ninety days will shape how the global technology and financial establishment thinks about AI-driven cyber threats. The frameworks, the terminology, the severity assessments — all of it will be calibrated by organizations that have had direct access to the model. We will receive a summary. A press release. Perhaps a whitepaper sanitised for public consumption.

Second-hand knowledge, in the language of cybersecurity, is the equivalent of knowing that a lock has been compromised after someone has already walked through the door.

India is not a peripheral player here. We are, by multiple credible assessments, among the most targeted nations for cyberattacks — financial institutions and banking infrastructure being the most vulnerable, followed by government systems and healthcare. This is not a hypothetical risk. The CloudSEK 2024 report and Palo Alto's Unit42 findings both pointed to this with uncomfortable specificity. And now we are being asked to prepare for AIaugmented attacks without having seen what those attacks might look like.

The governance response domestically has been moving in the right direction, though perhaps not at the pace events now demand. MeitY's AI Governance Guidelines, released late last year, establish sensible principles — a lighttouch, risk-based approach anchored in accountability, safety, and human oversight. The proposed AI Governance Group and the IndiaAI Safety Institute are welcome institutional steps. But structures take time to operationalise, and the interval between a framework being notified and it actually functioning is where the real vulnerability lives.

What is needed immediately is less structural and more diplomatic. The AI Leadership Summit earlier this year created exactly the kind of channel that should now be activated in outreach. Anthropic has recently appointed an India Managing Director — that relationship needs to be leveraged without ceremony or delay. The ask is modest: include two or three Indian organisations from high-risk sectors in Project Glasswing, or at minimum, arrange for a structured, non-public briefing for CERT-In and the relevant sectoral regulators before the public report drops.

It is also entirely plausible that the Mythos episode may be a calculated hype exercise — a move to generate urgency the way early AI companies sometimes manufactured alarm to accelerate adoption. But when the Fed Chair moves, it is usually not theatre.

Beyond the immediate, there is a structural lesson here. The pace of frontier AI development is now such that consequential capability shifts occur at monthly intervals, sometimes faster. We cannot staff a policy response to that pace if we are relying only on official company blogs and academic papers filtered through two news cycles. A practitioner-led task force — people who actually understand the models, who follow the technical discourse, who read the system cards — needs to be constituted and given real authority, not merely advisory status.

The asymmetry in information is not a permanent condition. It is, however, a policy choice. We can choose to close it or we can choose, through inaction, to widen it. The countries that will navigate this decade well are the ones that treat AI capability shifts not as technological curiosities but as matters of strategic consequence — because increasingly, they are indistinguishable from each other.