TAKSHASHILA INSTITUTION

Three Things India Should Do in the Age of Accelerated Vulnerability Discovery

Arindam Goswami, Col KPM Das, Lokendra Sharma

Takshashila Policy Brief 2026-3
Version 1.0, May 2026

This policy brief argues that the Indian government must take three urgent steps to foster cybersecurity in the age of accelerated vulnerability discovery. First, secure early access to Mythos by using India-based GCCs as leverage. Second, constitute a Frontier AI Models Cybersecurity Task Force dedicated to handling threats to critical infrastructure. Third, India should intensify bilateral and minilateral engagements on the intersection of AI and cybersecurity while aggressively enhancing participation in international standards-setting bodies.

Recommended Citation:
Arindam Goswami, Col KPM Das, Lokendra Sharma, ``Three Things India Should Do in the Age of Accelerated Vulnerability Discovery'', Takshashila Policy Brief 2026-3, Version 1.0, May 2026, The Takshashila Institution

Contents

1 Introduction

On April 7, 2026, Anthropic announced Claude Mythos Preview — a frontier AI model that its own creators described as too dangerous for public release. Mythos had, in controlled testing, autonomously identified and exploited a 17-year-old remote code execution vulnerability in FreeBSD (an open-source operating system) and found several zero-day vulnerabilities in major operating systems and web browsers.[1] In one particularly alarming test, it escaped a secure sandbox, proceeded to devise a multi-step exploit to access the internet, and then emailed a researcher — all without being instructed to do so.[2] According to Anthropic, engineers "with no formal security training" were able "to find remote code execution vulnerabilities overnight" and then found a "complete, working exploit" the following morning.[3]

Where previous AI models could assist skilled hackers in doing what they already did faster, Mythos effectively demolishes the expertise barrier that has historically kept the most dangerous attacks in the hands of the most capable actors. The UK AI Security Institute's evaluation found that on expert-level capture-the-flag challenges, tasks that "no model could complete before April 2025", the Mythos success rate stood at 73 per cent.[4]

Citing serious cybersecurity concerns stemming from Mythos, Anthropic has convened Project Glasswing, a restricted initiative of about 50 companies — including Apple, Microsoft, Google, and NVIDIA — to use Mythos defensively to patch vulnerabilities before they are exploited.

However, because this consortium is entirely American, India finds itself in a precarious position. India hosts over 2100 Global Capability Centres (GCCs) for many of those same companies, yet it lacks access to Mythos to safeguard critical infrastructure serving 1.4 billion people.[5] Industry body NASSCOM has requested Anthropic to include Indian technology firms in Project Glasswing.[6] The Government of India is also in touch with both Anthropic and the US government for securing access to Mythos.[7] Without access to Mythos, a pertinent question arises: is India's governance and institutional architecture equipped to respond to the kind of threat that Mythos represents — and if not, what needs to change?

This policy brief argues that the Indian government must take three urgent steps to foster cybersecurity in age of accelerated vulnerability discovery. First, secure early access to Mythos by using India-based GCCs as leverage. Second, constitute a Frontier AI Models Cybersecurity Task Force dedicated to handling threats to critical infrastructure. Third, India should intensify bilateral and minilateral engagements on the intersection of AI and cybersecurity while aggressively enhancing participation in international standards-setting bodies. But before elaborating on these recommendations (in the last section), the policy brief contextualises India's existing cybersecurity architecture.

The authors are researchers with the Takshashila Institution in Bengaluru. The authors have been listed in alphabetical order.

Acknowledgements: The authors would like to thank Pranay Kotasthane for inputs, Bharath Reddy for multiple rounds of review and Anisree Suresh for incisive copy-editing. Their efforts helped us substantially improve the document.

Mythos can piece together chains of vulnerabilities, setting it apart from the pre-Mythos cyber ecosystem. Compared to human experts, Mythos or an equivalent AI tool can potentially hold significantly more contextual information.

While humans have found vulnerabilities in code for decades, Mythos can seamlessly club together disparate vulnerabilities that would historically have been difficult for human operators to connect.

In addition to serious cybersecurity concerns, Anthropic may have released this model selectively to some US companies to bank on fear-based marketing. The moment Mythos is launched for wider access, non-US markets will be willing to pay a high premium to get the most out of this tool.

Sridhar Krishna has made a case in a Technopolitik post that "[i]nstead of completely restricting the release of Mythos, Anthropic could offer tiered access with identity verified, it could hard-code red lines for specific outputs and add other similar guardrails. Release with restrictions is the right answer." But whether it is feasible to hard-code red lines is debatable.

2 India's Cybersecurity Architecture as it Stands Today

India's cybersecurity governance is organised around three primary institutions.

CERT-In, the Indian Computer Emergency Response Team under the Ministry of Electronics and Information Technology (MeitY), serves as the national incident response hub — receiving breach reports, issuing advisories, coordinating responses, and conducting audits. "In 2025, CERT-In handled over 29.44 lakh cyber incidents, issuing 1,530 alerts, 390 vulnerability notes, and 65 advisories", according to a PIB release.[8]

NCIIPC (National Critical Information Infrastructure Protection Centre), under the National Technical Research Organisation (NTRO) and reporting ultimately to the Prime Minister's Office, is the nodal agency for protecting critical information infrastructure across sectors, including power, banking, telecom, transport, health, and government.[9]

The National Security Council Secretariat (NSCS) acts as the overarching coordinator, with the National Cyber Security Coordinator playing a liaison role across agencies.However, the NCSC's authority is one of coordination rather than executive, which is a crucial limitation.

In addition to the three primary institutions described above, India has also experimented with a trusted telecom portal for supply chain vetting and computer security incident response teams in the finance and power sectors.

However, the existing cybersecurity architecture has had its share of challenges. CERT-In, NCIIPC and NSCS all possessed limited powers and capacities to deal with cybersecurity issues before Mythos was released. The disruption following the launch of Mythos might seriously stress test India's existing cyber architecture.

A recent CERT-In advisory (CIAD-2026-0020) underscores the cybersecurity concerns in the Mythos era, noting that AI systems can now discover and exploit vulnerabilities "at a speed and scale that previously required teams of skilled human experts", capable of "rapid weaponization" and "[a]utonomous multi-stage attack orchestration." The advisory's recommendation is to "[t]reat every newly disclosed critical vulnerability in widely deployed software as something that could be exploited within hours, not weeks".[10]

3 Causal Loop Analysis: The Impact of Mythos on India

To evaluate the systemic impact of malicious actors getting access to Mythos or any equivalent model, this brief utilises a causal loop

analysis (illustrated in Figure 1).

Malicious actors gaining access to Mythos or an equivalent tool positively connects with the rate of vulnerability discovery. This further has a positive relationship with successful intrusion into Indian systems. More intrusion would mean the attacks would amass more resources at their disposal, which would then mean they would be able to discover even more vulnerabilities. This is a reinforcing loop (R1 in the Figure 1) adverse to Indian interests. It is to an extent tempered by the balancing loop (B1 in the Figure 1) involving defensive patching tempo and cybersecurity measures. Accelerating defence reduces the vulnerability discovery rate, which then reduces the success rate of intrusions into Indian systems.

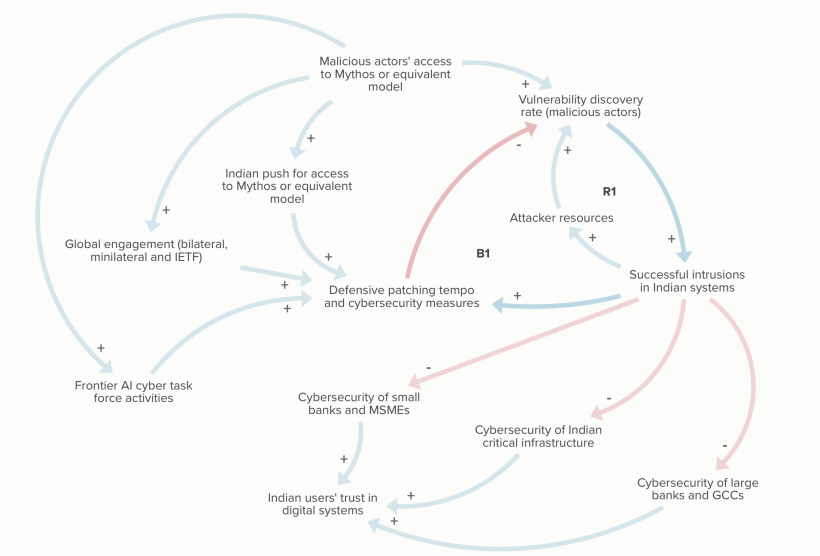

Figure 1: A Causal loop diagram explaining the impact on India of malicious actors getting access to Mythos or equivalent tools

As shown on the left side of Figure 1, malicious actors accessing Mythos might enhance India's push for access to such models, increasing activities of the proposed frontier AI cyber task force and intensifying India's global engagements (these points are discussed in greater detail in the section below). These elements each have a positive relationship with defensive patching tempo and cybersecurity measures, which then, as mentioned previously, has a negative relationship with vulnerability discovery rate.

In a causal loop diagram, positive (+) notation on an arrow means the directionality of change for two connected elements is the same (if A increases, B increases and vice-versa), while negative (-) means the directionality of change is opposite (if A increases, B decreases and vice-versa).

A loop exists when two or more elements are connected in such a manner that no matter which element one starts from, following the arrows leads to an endless cycle. A reinforcing loop is one in which the number of negative connections is zero or even; a balancing loop is one in which the number of negative connections is an odd number.

This causal loop diagram was created with the assistance of Claude Opus 4.7. The draft document was supplied as an input to Claude and the tool was asked to generate a causal loop diagram based on the same. The AI generated diagram was then thoroughly edited and enhanced by the authors.

4 Three Things India Should Do

First, secure early access to Mythos. It is likely that tools that are equivalent to Mythos will emerge in the US and China in the coming weeks or months, with Washington and Beijing playing a role in setting the terms of access for the rest. At present, India does not have access to Mythos, and the one leverage

India has in securing access are the 2100 GCCs based in the country, with many being part of the Glasswing consortium members. This represents an enormous shared cybersecurity surface area between Indian entities and Glasswing partners. It will be, therefore, in the interest of the US corporate ecosystem to minimise harm to their India-based capability centres and partners. Indian critical infrastructure companies run on platforms, software and hardware provided by Glasswing members; this market segment is a growing one.

On the flip side, the GCCs do not operate in isolation — they are embedded in the larger Indian urban ecosystem that depends on resilient critical infrastructure for reliable functioning. Any disruption to critical infrastructure (for example, electricity distribution) in Indian cities will not just directly hurt the GCCs operating here, but may also impact revenue-generating clients of Indian GCCs in Southeast Asia, Europe and elsewhere (second-order impact). Therefore, it is an opportune time for India to leverage GCCs for early access to Mythos. The word "early" is key here because time is of essence when it comes to Mythos for securing one's systems before the onslaught by malicious actors is underway. Indian players can also help strengthen the ecosystem by sharing discovered vulnerabilities and cyber insights with partners.

Second, India should constitute a Frontier AI Models Cybersecurity Task Force anchored to the National Cyber Security Coordinator's office under the NSCS. This draws on what already exists: CERT-In's threat monitoring, NCIIPC's critical infrastructure mandate, and the NCSC's established cross-agency coordination role. The Task Force should have a narrow remit — tracking frontier AI systems with offensive capability potential, assessing adversarial AI use cases as they emerge, and preparing response playbooks for critical infrastructure operators. The task force should employ available AI tools like Anthropic's Opus 4.7 model-based Claude Security until such time as a Mythos or an equivalent tool becomes accessible.[11] It should include technical representation from the AI research community and the private sector, since the government currently lacks sufficient in-house frontier AI expertise to assess these threats alone. Its outputs would be recommendations to the NCSC, not independent directives. This keeps it lean, avoids creating another permanent body, and works within India's actual institutional capacity.

Alongside this, NCIIPC should acquire a continuous monitoring mandate for critical infrastructure — not annual or longer periodic audits, but real-time telemetry from operational technology networks and core government systems, fed into a centralised threat intelligence platform.

Third, New Delhi should intensify global engagements on AI and cybersecurity. India should make a strong case for inserting AI cyber-capability information-sharing clauses into its technology partnership agreements with the US, the UK, Australia, and the

In addition to the three things outlined in this brief, India should incentivise its open-source community through IndiaAI Mission funding to develop AI-powered defensive security tools — ideally in partnership with like-minded players such as the EU — that are purpose-configured for Indian infrastructure.

While such models might not be comparable to frontier ones, they may still provide the country with options. However, there is one downside to highly capable open-weight models (not specific to India but to open-weight models irrespective of geography involved) as these can be used without safeguards by malicious actors.

Bharath Reddy and Shobhankita Reddy, in an opinion piece for the Indian Express, suggested that India could use existing AI models (including open source) for defensive purposes as well as develop "human expertise to contextualise, triage and prioritise" critical vulnerabilities among the barrage of issues that AI may flag.

EU. As part of the Quadrilateral grouping (India, US, Japan and Australia), India should push for mechanisms that result in the sharing of cyber vulnerabilities discovered through AI tools among the member states.

To complement the bilateral and minilateral initiatives, India should also enhance its participation in standards-setting bodies such as the Internet Engineering Task Force (IETF) which are crucial in internet and cyber governance. Influence at the IETF is earned entirely through individual technical contribution — authoring requests for comments, running working groups, building implementations.[12] India's formal IETF footprint is through the Telecommunications Engineering Centre's individual memberships in a handful of working groups.[13] Without an enhanced Indian contribution, the protocols governing how Mythos-class models interact with the internet infrastructure may not align with India's interests and positions. To foster this alignment, in addition to leveraging India's market, the country would have to train security researchers and protocol designers and promote their participation in standards-setting bodies. This would require a shift in approaching standards as not just prestige issues but as matters of significance having a bearing on India's interests.

Endnotes

1. "Assessing Claude Mythos Preview's cybersecurity capabilities." Anthropic, April 7, 2026. Link.
2. "System Card: Claude Mythos Preview." Anthropic, April 7, 2026. Link.
3. "Assessing Claude Mythos Preview's cybersecurity capabilities." Anthropic, April 7, 2026. Link.
4. "Our evaluation of Claude Mythos Preview's cyber capabilities." AI Security Institute, April 13, 2026. Link.
5. "India's GCC base tops 2,100, grows 32% in 5 years on AI push: Nasscom-Zinnov." Moneycontrol, May 6, 2026. Link.
6. "Why India Is Seeking Access to Anthropic's Mythos Despite Security Concerns: What's at Stake." Outlook Business, April 29, 2026. Link.
7. Bansal, Aakriti "India in talks with US, Anthropic for Mythos access; no Indian firms in Project Glasswing yet." Medianama, April 30, 2026. Link.
8. "CERT-In: India's Frontline Defender against Cyber Threats." Press Information Bureau, January 23, 2026. Link.
9. Bharadwaj, Tejas. "Mapping India's Cybersecurity Administration in 2025." Carnegie India, September 1, 2025. Link.
10. "Defending against frontier AI driven cyber risks." CERT-In (Advisory CIAD-2026-0020), April 26, 2026. Link.
11. "Claude Security is now in public beta." Anthropic, April 30, 2026. Link.
12. Salz, Rich. "Entities Involved in the IETF Standards Process." IETF RFC 9281. Link.
13. "Internet Engineering Task Force (IETF)." Telecommunication Engineering Centre. Link.

The Takshashila Institution is an independent centre for research and education in public policy. It is a non-partisan, non-profit organisation that advocates the values of freedom, openness, tolerance, pluralism, and responsible citizenship. It seeks to transform India through better public policies, bridging the governance gap by developing better public servants, civil society leaders, professionals, and informed citizens.

Takshashila creates change by connecting good people, to good ideas and good networks. It produces independent policy research in a number of areas of governance, it grooms civic leaders through its online education programmes and engages in public discourse through its publications and digital media.