MANOHAR PARRIKAR INSTITUTE FOR DEFENCE STUDIES AND ANALYSES
मनोहर पर्रिकर रक्षा अध्ययन एवं विश्लेषण संस्थान

MP-IDSA
Issue Brief

The Making of a Mythos: AI-Enabled Cybersecurity and the Emerging Architecture of Access

Meghna Pradhan

May 20, 2026

Summary

Anthropic's release of the frontier AI model Claude Mythos Preview has caused significant trepidation, as its autonomous code-reasoning capabilities surfaced previously undetected vulnerabilities across major operating systems and software infrastructure. The architecture of access, being constructed around such capabilities, creates an asymmetry in access. India needs to secure critical infrastructure by updating, modernising or replacing legacy systems.

Introduction

In April 2026, Anthropic stated that its frontier artificial intelligence model, dubbed Claude Mythos Preview, had unprecedented cyber capabilities that made it too dangerous to release publicly.[1] They disclosed that the Large Language Model (LLM), during its testing phase, had identified 'thousands of high and critical severity vulnerabilities', many of which have been, and continue to be, validated by cybersecurity experts. These include discovering and exploiting a 27-year-old Zero Day Vulnerability (0-day) in OpenBSD, finding a 16-year-old bug in media processing library FFMpeg, and patching 271 vulnerabilities in Firefox 150 based on a single evaluation cycle.

While Mythos Preview's capabilities mark a significant step in the development of AI systems, their potential has rung alarm bells among global cybersecurity communities, financial institutions and governments. The global financial elites have especially highlighted cyber threats posed by Mythos. Anthropic has disclosed that Mythos was used to conduct what could essentially be called 'digital robbery' on systems across the world, which has led to fears that the system may further be misused for mass looting of bank accounts,[2] paralysis in international payment systems, or spark a crisis of confidence in extant financial systems. [3]

Global finance leaders and central bankers from countries such as the US, Canada, the EU, the UK, Germany, South Korea and India have expressed concerns and interest in accessing the LLM to plug vulnerabilities in their financial systems. Simultaneously, Anthropic has launched Project Glasswing, an AI consortium comprising companies such as Amazon Web Services, Apple, Broadcom, Cisco, CrowdStrike, Google, JPMorgan Chase, the Linux Foundation, Microsoft, NVIDIA, and Palo Alto Networks, as well as over 40 other organisations with access to the system.[4]

The discussions around Mythos and its capabilities are being presented as a singular new threat, a framing that does hold some merit, but does not present a complete picture. Mythos did not create the vulnerabilities it found and developed exploits for; zero-day vulnerabilities have always existed and have been routinely exploited. What has changed in Mythos is the tempo of these threats, as it accelerates the

[1] Joe Tidy and Imran Rehman-Jones, **"Anthropic Investigating Claim of Unauthorised Access to Mythos AI Tool"**, *BBC*, 22 April 2026.

[2] Margi Murphy, Jake Bleiberg and Patrick Howell O'Neill, **"How Anthropic Learned Mythos Was Too Dangerous for the Wild"**, *Bloomberg*, 16 April 2026.

[3] Laura Noonan, Colum Murphy and Jorge Valero, **"Mythos AI Sparks Fear and Confusion Among Global Finance Elite"**, *Bloomberg*, 16 April 2026.

[4] Matt St. Jean, **"Project Glasswing Found Thousands of Zero-Days. Closing It Requires More Than a Faster Patch Cycle"**, *Blue Mantis*, 28 April 2026.

weaponisation of extant fragilities. It is therefore crucial to examine how the capability jump, cost dynamics, human oversight, access distribution, and structural issues interact with the cyber-capabilities Mythos has heralded.

Contextualising the Global Furore

The case of OpenBSD has effectively illustrated the reason for the alarm at Mythos. The flaw had remained undetected in one of the 'most security-hardened operating systems ever built'[5] since around 1998, despite hundreds of code revisions, security reviews and audits. The delay in discovery may be due to multiple factors: it required a very specific sequence of reasoning among thousands of possible lines of reasoning and permutations. A human auditor may miss it, not due to incompetence, but because the sheer volume of code and the complexity of interactions between components may tax the limits of human cognition. Additionally, programmers often triage the code based on perceived criticality and personal experience, thereby prioritising newer code over stable, mature codebases. Mythos has no compunctions about dealing with legacy code.[6]

Mythos have added considerable dimension to the potential of AI in cybersecurity. While current AI systems do have a degree of autonomy, they require human guidance to reach the desired result. Mythos found over 271 vulnerabilities in Firefox in April 2026 without additional guidance after the initial command.[7] By comparison, Anthropic's Claude Opus 4.6 found only 22 bugs in a similar study, all of which required human steering. The more consequential capability shift, however, is how Mythos's autonomy applies to what it finds. Unlike current frontier models, Mythos not only identifies vulnerabilities but can also independently act to close the loop from code analysis to the development of working exploit chains, with minimal human involvement.

Typically, finding a bug and developing an exploit are distinct stages, and the time and capability required to bridge the gap give defenders a window to patch before weaponisation. Mythos, however, collapses this window to the same timescale as discovery itself.[8] The dual-use nature of AI technologies, such as Mythos, means that

[5] Morgan Ellis, **"An AI Found a 27-Year-Old Bug Hiding in OpenBSD. It Cost Less Than $50 to Find It"**, *The Medium*, 18 April 2026.

[6] Dominik Waitzer, **"AI Security Audits: Anthropic Mythos Discovers 27-Year-Old OpenBSD Vulnerability"**, *Desight Studio*, 8 April 2026.

[7] Lily Hay Newman, **"Mozilla Used Anthropic's Mythos to Find and Fix 271 Bugs in Firefox"**, *The Wired*, 21 April 2026.

[8] **"The Machine That Hunts Ghosts Claude Mythos, GPT-5.4-Cyber, and the AI Tools Reshaping Cybersecurity"**, *VARINDIA*, 10 May 2025.

the same capability that enables a defender to audit a codebase can also enable an attacker to weaponise its findings. Furthermore, reports suggest that during evaluations, Mythos appeared to know it was being tested, showing signs of this awareness in roughly 29 per cent of transcripts, even when it didn't explicitly say so. It may also have intentionally underperformed to avoid flagging behaviour that might seem disingenuous, thereby indicating a nascent capability to manage perception by manipulating responses.

Cost is the second vector to consider. The scan that found the OpenBSD vulnerability used less than US$ 50 in resources and was part of a broader 1,000-scaffold run that cost US$ 20,000.[9] Bounties for comparable bugs typically run much higher, requiring weeks or months of dedicated work. Cybersecurity experts also tend to be paid significantly higher salaries, based on their expertise and market dynamics. This has two major implications: one, AI systems like Mythos may be leveraged to find critical vulnerabilities at a fraction of human resources. Two, leveraging AI for vulnerability discovery may introduce predictability in cybersecurity costs.[10]

Broadly, the two factors above have significant implications for the human role in AI-enabled processes, particularly in cybersecurity. The potential for lower, more predictable costs due to access to AI capabilities (at least as good as, or comparable to, what Mythos claims to have), especially with higher-order autonomy, has implications for how much human capabilities can match the resulting collapse in timelines. As of now, there is still a need for human expertise in identifying and verifying vulnerabilities; Anthropic itself has launched a public bug bounty programme on the threat-exposure management company HackerOne, inviting external human researchers to identify vulnerabilities in its systems.[11]

However, there is a need to be aware of the possibility that rising autonomous AI capability to find a vulnerability and execute the result (patching and/or developing an exploit) autonomously, combined with LLM systems' emerging capability to predict and manipulate their responses, may not only weaken any meaningful assertion of human oversight but also significantly reduce the need and demand for human expertise.

---

[9] Morgan Ellis, **"An AI Found a 27-Year-Old Bug Hiding in OpenBSD. It Cost Less Than $50 to Find It"**, *The Medium*, 18 April 2026.

[10] The current market with human researchers has been priced based on capability, geography and supply, creating vastly different cost structures. If AI resources can be calculated and priced accordingly, the dynamism in market pricing driven by premium charged by cybersecurity experts may see a downward trend.

[11] Paul Sawers, **"Anthropic Puts the 'Myth' in Mythos with its HackerOne Bug Bounty Program"**, *The New Stack*, 10 May 2026.

The Differential Architecture of Access

Anthropic's response to Mythos's capabilities has been to create an AI consortium consisting of some of the major big tech companies and financial institutions. On the surface, the logic of Project Glasswing remains defensible: a public release of the LLM would give malicious actors the tools to damage structures that the current cybersecurity apparatus is not equipped to defend. However, the structure of this project also raises important concerns. Access to a system with such glaring implications for the security of global financial and cyber systems is being administered at the discretion of a private entity to other private entities, without a statutory framework, treaty obligations, or accountability mechanisms. Additionally, the selected partners are, barring a few exceptions, largely American companies, most of which already have access to advanced AI capabilities and strong cybersecurity.

There are additional reports that Anthropic has been engaging with US national-security institutions to negotiate terms of access to Mythos, possibly as part of a broader effort to rebuild relationships strained since the fallout from the earlier controversial defence AI partnership.[12] The US had initially restricted Anthropic from expanding the number of entities for strategic reasons.[13] The US Commerce Department's Center for AI Standards and Innovation, along with Anthropic, OpenAI, and Google, is also discussing the potential development of a pre-deployment evaluation framework.[14]

The geopolitical implications of Mythos extend beyond Anthropic's provision of access. In Hong Kong, Goldman Sachs removed access to all Claude models in late April.[15] The timing of the banking entity's 'strictest interpretation of its contract' is suggestive of Mythos-related cybersecurity concerns.[16] There is an emerging pattern of the concentration of AI capabilities and the restriction of access along lines of commercial interest and political logic, even as the prominent narrative centres on taking responsibility for a powerful yet neutral commercial product.

---

[12]Casey Newton, **"Why Anthropic's New Model Has Cybersecurity Experts Rattled"**, *Platformer*, 7 April 2026.

[13] Primary possible reason was US concern regarding sufficiency of compute with Claude for government's use; more entities will reduce the amount of compute available per partner; Robert Mcmillan, **"White House Opposes Anthropic's Plan to Expand Access to Mythos Model"**, *The Wall Street Journal*, 30 April 2026.

[14] **"Trump Administration to Test New AI Models from Google, Microsoft, xAI Before Public Release"**, *Hindustan Times*, 6 May 2026.

[15] While ChatGPT and Claude were never officially supported in Hong Kong, companies did have access to their tools.

[16] With specific reference to China, both ChatGPT and Claude remain banned. The specific restrictions from Goldman Sachs also extend to any of their overseas employees visiting the Hong Kong office, indicating complete shutdown of the AI LLM. **"Goldman Cuts Access to Anthropic's Claude for Hong Kong Bankers, Source Says"**, *Reuters*, 29 April 2026.

The case for responsible disclosure and meaningful safety was also weakened by a leak in the access architecture that occurred on the day of the Glasswing project's announcement. A small group of users gained unauthorised access to Mythos through their third-party vendor, Mercor.[17] Given Mythos's capability to collapse discovery-to-exploit timelines, unauthorised access carries serious implications on a global scale. There was no legitimate accountability for the unauthorised access, leaving scope for adversarial actors to attack a system at machine pace without credible deterrence. At the same time, entities without comparable capabilities will be rendered incapable of effective response.

Developments in other frontier models also indicate that the cyber-capability gap among AI-cyber capabilities may not be insurmountably vast. Research by the UK's AI Security Institute (AISI), conducted through a series of Capture-The-Flag challenges, suggested that OpenAI's GPT-5.5 has also shown cyber capabilities comparable to Mythos.[18] Similarly, in China, Qihoo 360 won the 2026 Tianfu Cup hacking competition by leveraging AI to develop a 'Vulnerability Discovery Agent', which reportedly identified around 1,000 software vulnerabilities at a pace comparable to Mythos.[19] The Chinese example is especially critical: while Mythos tests were held in a controlled sandbox environment where cybersecurity exigencies were disabled, Qihoo's runs were conducted in real-world adversarial competition against production environments, making the latter's use more closely reflect how a cybersecurity attack will unfold in the wild.

The rhetoric of restricted access to capability may also be redundant if the results of AI systems are repeatable and/or exploitable despite guardrails placed to prevent it. Independent laboratories such as Vidoc Security Lab[20] and AISLE[21] were able to use cheaper, publicly available models such as GPT-5.4 and Claude Opus 4.6 to reproduce findings similar to Mythos (though without the exact accuracy). These

[17] Mercor, an AI feedback recruitment company, was hacked using another third-party tool called LiteLLM. Since Anthropic does not vet cybersecurity measures of all its third-party vendors, the original hack that had already subjected Mercor to a major data breach, also opened doors for use of Mythos. Jon Martindale, **"How a Cavalcade of Blunders Gave Unauthorized Users Access to Claude Mythos — Restricted Model Accessed by Third Parties, Thanks to Knowledge from Data Breach"**, *Tom's Hardware*, 4 May 2026.

[18] GPT 5.5 is as of now available to paid users, and its more advanced cybersecurity features are under a tightly controlled access programme for 'critical cyber defenders' in sectors like energy, finance and key digital services. See Shalabh Singh, **"GPT-5.5 Cyber Breakthrough: Powerful New AI Shields Critical Systems"**, *Pune Mirror*, 1 May 2026.

[19] Piyush Shukla, **"Latest Shot on Anthropic's Mythos—China's Cybersecurity Giant Qihoo 360 Finds 1,000 Software Vulnerabilities Fast, Raising Global Zero-day Risks"**, *The Economic Times*, 22 April 2026.

[20] Dawid Moczadło, Klaudia Kloc, Marek Lewandowski, Amadeusz Lisiecki, Jakub Sienkiewicz and Mikołaj Palkiewicz, **"We Reproduced Anthropic's Mythos Findings With Public Models"**, *VIDOC*, 14 April 2026.

[21] Stanislav Fort, **"AI Cybersecurity After Mythos: The Jagged Frontier"**, *AISLE*, 7 April 2026.

experiments suggest that older models can find critical vulnerabilities at comparable rates, since most of the raw intelligence needed for it is generally available.[22] Additionally, jailbreaks[23] remain a regulatory grey area and may be leveraged to bypass guardrails on even the most advanced frontier models. This is a concern, as jailbreak techniques persist despite repeated patching by AI companies.

Essentially, access frameworks to models like Mythos show a far more chequered outlay than its developers tend to present. Transparency is a relative currency, subordinate to geopolitical and commercial interests. While there may be some claim to ethical and responsible insurance against malicious use, the lack of formal mechanisms to enforce them makes advanced capabilities reachable both legally and illegally. While companies may impose guardrails to restrict some channels of use, no model can restrict them all.

India in the Age of Mythos: Predicaments and Hard Choices

The Mythos reveal has sent alarm bells ringing across the globe, and India is no exception. On 23 April 2026, Finance Minister Nirmala Sitharaman convened a high-level meeting. Minister of Electronics and Information Technology Ashwini Vaishnaw, heads of banking and key cybersecurity bodies assessed Mythos-related cybersecurity risks to financial institutions.[24] While deeming Mythos an unprecedented challenge to the Indian financial system, the Finance Minister instructed banks to secure their IT infrastructure and protect customer data, and emphasised the need for a system that allows real-time sharing of threat intelligence among banks and cybersecurity agencies.[25] Additionally, a panel has been set up under State Bank of India Chairman C.S. Setty to assess risks arising from the AI platform Mythos and develop mitigating measures.[26]

India is not alone in recognising Mythos as a national security concern; there has been a broader trend of countries viewing the emergent capability with some trepidation. India does face unique constraints in any decision regarding access to

[22] The main argument is that Mythos does not bring anything new in 'finding vulnerabilities' aspect of cybersecurity. The game has changed in terms of managing and validating them autonomously. See Dawid Moczadło, Klaudia Kloc, Marek Lewandowski, Amadeusz Lisiecki, Jakub Sienkiewicz and Mikołaj Palkiewicz, **"We Reproduced Anthropic's Mythos Findings With Public Models"**, no. 20.

[23] Jailbreak refers to techniques that enable bypassing safeguards and restrictions manufacturers and service providers place on their products to prevent misuse.

[24] **"Nirmala Sitharaman Urges Bankers to Brace for AI Threats Amid Concerns Over Anthropic's Mythos"**, *The Hindu*, 23 April 2026.

[25] **"Claude Mythos Threat: Nirmala Sitharaman Flags 'Unprecedented' AI Cyber Risk, Banks on High Alert"**, *PTC News*, 24 April 2026.

[26] **"Govt Forms Panel Under SBI Chief C S Setty to Assess Mythos-related Risks"**, *The Economic Times*, 24 April 2026.

advanced AI cybersecurity capabilities or protection against them. The issue of access has been at the forefront for Indian entities such as the National Payments Corporation of India (NPCI) and Paytm, as evidenced by their engagement with Anthropic and the US Government to secure early access. However, the push for data localisation and sovereign data within India means that auditing critical sector entities, such as NPCI's systems, with a completely foreign-based AI system may be counterintuitive.[27] The US decision to prevent Anthropic from including additional entities in Project Glasswing adds another hurdle to India's bid for access.

The Indian banking system also faces structural issues that predate Mythos. India has been pursuing last-mile connectivity, financial inclusion, and the digitisation of governance structures, bringing millions of Indians into the fold of formal banking through the JAM trinity and the Unified Payments Interface (UPI). However, such large-scale programmes have come at a cost: India's Core Banking System (CBS) has been supported by centralised technological structures that are notoriously slow-moving and difficult to change.[28]

Therefore, banks in India develop specialised software stacks around CBS to meet specific business demands, adding multiple layers to it.[29] There are two major consequences of this: first, any vulnerabilities discovered within the system would likely have grown more complex and would require longer timelines to fix. Second, legacy systems also come with legacy zero-day vulnerabilities that may have evaded cyber-audits and remain susceptible to exploitation.

India also has a record of a relatively sluggish pace in patching vulnerabilities, even in critical systems. Wannacry Attacks had proven to be an illuminating example: the Windows vulnerability exploited by the ransomware had been discovered and patched by Microsoft two months before the attack. However, the affected Indian systems were either using outdated software or were not patched to the endpoint.[30] Similarly, the 2022 AIIMS ransomware attack has been strongly attributed to the use of outdated systems and/or systems with known vulnerabilities.[31] Indian power grids remain inadequately protected despite 2021 cybersecurity guidelines, largely due to resource constraints.[32]

---

[27] Arghanshu Bose, **"What Global Banks Said About Anthropic's New Mythos Model That the Company 'Refused' to Release Publicly"**, *The Times of India*, 27 April 2026.

[28] **"What Software is Used in the Banking Sector in India"**, *IGCB*, 12 December 2025.

[29] **"Mythos Shrinks Exploit Time; Indian Finserv Patching Lags"**, *Let's Data Science*, 4 May 2026.

[30] **"WannaCry Impact on India Under-reported"**, *The Hindu*, 17 November 2017.

[31] Adarsh Nautiyal, **"Cyber Resilience in Healthcare Lessons from AIIMS Delhi's Cybersecurity Battles"**, *Pivot*, 3 March 2024.

[32] Aggam Walia, **"Amid Rising Cyber Threats, Power Grid Told Officials Over 270 Substations Lack 'Next-generation' Firewalls"**. *The Indian Express*, 16 May 2025.

The aforementioned case studies are representative of deeply rooted, structurally compounded issues in Indian cybersecurity. Access to Mythos alone will not resolve persistent challenges around patch transmission, cybersecurity capacity, and legacy system failures. Most critical-sector entities and cybersecurity bodies in India operate in resource-constrained environments, and the volume of high-severity findings a Mythos-class system will unearth may overwhelm them. India's cyber threat profile is also evolving quickly. Countries that have featured prominently in attribution reports of cyberattacks against Indian infrastructure, including China, have been rapidly developing AI-enabled cyber capabilities. To keep up, India's response will need to be calibrated to the accelerating tempo and scope of the threat.

Within these constraints, India's response space is structured by three principal options, each with characteristic costs.

- **Full purchase** of Mythos-class and any potential Mythos-audited software systems is one of the seemingly obvious and most disadvantageous options for India. While it offers the shortest path to operational deployment, it is extremely expensive. This also increases India's dependency in the domain. Lower-resource buyers are typically offered older versions of commercial software, with extended security support running on time-limited contracts that may require renegotiations in cost and/or conditions. This option is also most vulnerable to supply chain risks and political control, as the slightest shifts in geopolitics may lead to the revocation of access. Purchase therefore secures access only for as long as the seller continues to sell, on terms the seller continues to find commercially acceptable.
- **Collaboration** is another option, specifically with companies that are already part of the Glasswing project. Major Indian IT services firms, such as TCS and Infosys, already serve as strategic partners to Glasswing entities, including Microsoft and Google. They could plausibly serve as channels through which Mythos-derived defensive findings flow to Indian critical infrastructure. This route preserves partial control for India while enabling some transfer of expertise. However, in such circumstances, negotiation asymmetry becomes a significant issue: foreign partners typically retain high-value intellectual property and may constrain the flow of capabilities in favour of confidentiality. Such negotiations also remain subject to the partnering entity's geopolitical priorities, as evidenced by US restrictions on Glasswing's expansion.

- **Indigenisation** of critical software systems offers sovereign control over systems and the pacing of their updates, which can be synergised against the threat landscape for India. However, the major hindrance to this remains feasibility. Most critical systems in the country, including processes and hardware, are largely dependent on a foreign software stack. MayaOS has proven to be a critical example here: the Ubuntu-based operating system was developed by the Defence Ministry from 2021 to replace Windows in security-sensitive government systems.[33] Yet, shifting the workflow from the Microsoft-enabled system to MayaOS has made slow progress, with integration largely limited to a few Ministry of Defence systems. With accelerating timelines driven by AI-enabled cyber threats, slow progress in indigenous systems integration may not be ideal, even if such systems offer greater protection. The option may be feasible on its own in the medium to long term, but it will require far more investment and effort to implement in the protracted threat timelines that systems like Mythos indicate.
- **Sovereign Deployment** of AI models may strike a balance in the current scenario. As seen in the examples from Vidoc and AISLE, lower-end models can be oriented to achieve results that closely parallel those of Mythos. India can leverage indigenous AI efforts such as BharatGen, Sarvam, and the IndiaAI Mission to build security assets specific to its needs. While they may lag in raw compute and cybersecurity capabilities in the near term, it may help India develop the necessary scaffolding to grow its cybersecurity assets. For aspects where the indigenous model may not suffice, open models available internationally may serve as a stopgap.

Whichever one (or mix) of these choices India opts for, getting hold of better tools is only half the answer. Indian institutions also need to make better use of the tools they already have. A scan that flags a vulnerability is only useful if the patch reaches the actual machine. Critical institutions often lack full awareness of their software systems and the personnel to respond quickly when issues are flagged. The use of legacy systems, especially when their operators do not always issue security updates, also poses a major vulnerability. Instead of focusing dialogue solely on Mythos, it is necessary to invest in and strengthen cybersecurity capacity for critical sectors, particularly by building a clearer view of which software those institutions run versus what they need, and gradually retiring older systems.

[33] John Xavier, **"Why is India's Defence Ministry Ditching Microsoft Windows for Ubuntu-based Maya OS?"**, *The Hindu*, 14 August 2023.

Conclusion

The dialogue around Mythos has largely shaped it as a threat, and while it has real potential for harm, its capability can also be viewed as a forcing function. After all, Mythos was not designed as a cybersecurity tool. Anthropic built it as a general-purpose model with extended autonomous reasoning capabilities across scientific research, legal analysis, strategic synthesis, and complex systems modelling. The cybersecurity capability surfaced during the evaluation of those reasoning capabilities, and Anthropic chose to highlight it in its subsequent disclosures. The infamy Mythos has acquired is therefore not a product of deliberate design intent, but of inadvertent capabilities the model came to exhibit. This is a critical inflexion point. As AI systems become more capable, any sufficiently developed system, regardless of the purpose for which it was built, may develop properties that its designers did not intend. Still, the wider world is left to confront and absorb.

The threat perception of Mythos is intensified by the architecture of access surrounding it. Unlike its use case, Mythos's narrow access structure through Project Glasswing was a deliberate choice. A private corporation has assumed significant gatekeeping authority over a capability with implications for most critical systems globally (including Indian entities), and any institution that lacks access to it risks asymmetric exposure and little defence. It is crucial, especially for countries that face similar exposure due to asymmetry in AI capabilities, to coordinate and press for transparent international norms on how systems are tested, and to build the capacity to verify capability claims themselves.

For India, Mythos is not an issue because it creates a novel vulnerability, but because it significantly compresses the timeline for existing threats, even as the resources, infrastructure, and capability to cope with them remain lagging. India is at an unenviable crossroads with no easy or perfect choice; each path has a cost, and none guarantees immediate and/or durable security. Modernising current systems and accelerating patching timelines in India are as critical as pursuing frontier capabilities that do not force a choice between security and sovereignty. The emergence of Mythos, therefore, should not generate panic but rather evoke caution about the nascent stage of AI development, characterised by the dilution of human oversight, inequitable access to capabilities, and collapsing timelines.

About the Author

Ms. Meghna Pradhan is Research Analyst at the Manohar Parrikar Institute for Defence Studies and Analyses, New Delhi.

Manohar Parrikar Institute for Defence Studies and Analyses is a non-partisan, autonomous body dedicated to objective research and policy relevant studies on all aspects of defence and security. Its mission is to promote national and international security through the generation and dissemination of knowledge on defence and security-related issues.

Manohar Parrikar Institute for Defence Studies and Analyses
1, Development Enclave, Rao Tula Ram Marg
New Delhi 110 010 India
T +91-11-2671 7983 F +91-11-2615 4191
www.idsa.in
Twitter @IDSAIndia
www.facebook.com/ManoharParrikarInstituteforDefenceStudiesAnalyses