When AI Becomes the Attack Vector: Banking's New Security Challenge

Posted on May 11, 2026 by CISO Forum Bureau

Finance Minister Nirmala Sitharaman's recent meeting with bank leaders and Electronics & IT Minister Ashwini Vaishnaw signals that India's financial sector faces a security challenge unlike any previous threat. The concern centers on Anthropic's Claude Mythos model, an AI system with unprecedented ability to identify digital security vulnerabilities – capabilities that could defend systems when used legitimately or devastate them when weaponized by attackers.

Govind Rammurthy
CEO & Managing Director
eScan

According to Sitharaman's remarks to stakeholders, the threat from advanced AI models is "unprecedented and requires a very high degree of vigilance, preparedness and better coordination across financial institutions and banks." This assessment reflects a fundamental shift: cybersecurity threats are no longer just sophisticated attackers using advanced tools, but advanced tools themselves becoming autonomous threat actors capable of discovering vulnerabilities faster than human security teams can patch them.

The Claude Mythos model, announced April 7 as part of Anthropic's Project Glasswing, was designed for defensive cybersecurity – helping organizations identify weaknesses before attackers exploit them. The problem, as Finance Minister Sitharaman noted, is that "not much is known, not very many people have tested or tried" this technology, creating uncertainty about both its capabilities and how to defend against its potential misuse. When an AI system can analyze operating systems and discover vulnerabilities that human researchers might miss, the question isn't whether it will be weaponized it's when and how to prepare for that inevitability.

Indian banks make particularly attractive targets for AI-powered attacks for several converging reasons. First, rapid digitalization created massive attack surfaces as banking moved from branches to mobile apps, with security sometimes trailing functionality in priority. Second, the sheer transaction volume – UPI alone processes 15 billion monthly transactions – creates complexity where subtle vulnerabilities can hide amid legitimate activity. Third, India's banking infrastructure includes legacy systems running alongside modern platforms, creating integration points where security assumptions from different eras clash.

Traditional cybersecurity operated on relatively predictable timelines: attackers discover vulnerabilities, exploit them for weeks or months before detection, defenders analyze the attack, develop patches, and deploy fixes. This cycle, while reactive, at least operated at human speed where defenders could catch up eventually. AI-powered attacks collapse these timelines dramatically – vulnerability discovery, exploit development, and attack deployment potentially happening faster than human security teams can respond, even with automated tools.

The scale of Mythos' capabilities became starkly visible when Mozilla applied the AI model to Firefox 148, their stable browser version used by hundreds of millions globally. Claude Mythos discovered 271 vulnerabilities in a single evaluation pass – an unprecedented number that Mozilla's CTO Bobby Holley described as causing "vertigo" for his security team. "For a hardened target, just one such bug would have been red-alert in 2025, and so many at once makes you stop to wonder whether it's even possible to keep up," Holley wrote in Mozilla's blog post announcing Firefox 150's release with all 271 flaws patched.

To put this in perspective, when Mozilla had previously used Anthropic's earlier Claude Opus 4.6 model on the same codebase, it found just 22 security-sensitive bugs – Mythos discovered more than ten times that number. What's particularly concerning is Holley's admission that "computers were completely incapable of doing this a few months ago, and now they excel at it," with Mythos performing "every bit as capable" as elite human security researchers but at machine speed and scale. This isn't about finding exotic new vulnerability classes – it's about discovering massive numbers of exploitable flaws in mature, heavily audited code that human researchers and automated fuzzing tools had missed for years, fundamentally changing the timeline between vulnerability discovery and potential exploitation.

From eScan's perspective as a cybersecurity provider serving India's banking and financial sector for over 25 years, the AI threat requires rethinking security architectures rather than just adding AI-powered defenses to existing frameworks. Banks need endpoint detection and response systems that monitor for anomalous behavior patterns rather than just known malware signatures, because AI-generated attacks will likely use novel techniques that signature-based detection misses. Network security must analyze traffic patterns for subtle deviations that indicate reconnaissance or data exfiltration, even when individual transactions appear legitimate. Data loss prevention becomes critical because AI-powered attacks may focus on subtle, long-term data theft rather than obvious smash-and-grab ransomware.

The challenge extends beyond technical defenses to organizational preparedness. When attacks potentially happen faster than human response times, automated response capabilities become necessary – systems that can isolate compromised endpoints, block suspicious transactions, and contain threats without waiting for human authorization. This requires trust in automated systems that many organizations understandably hesitate to grant, because false positives in automated responses can disrupt legitimate operations as badly as actual attacks.

Minister Sitharaman's emphasis on "better coordination across financial institutions" recognizes that AI-powered threats don't respect organizational boundaries. An AI system discovering vulnerability in one bank's infrastructure likely found similar vulnerabilities across multiple banks using comparable technology. Information sharing about attempted attacks, successful defenses, and emerging threat patterns becomes critical infrastructure rather than competitive disadvantage, requiring industry-wide collaboration that hasn't historically characterized banking security practices.

The meeting's timing – occurring as India's banking sector processes record digital transaction volumes and expands financial inclusion to hundreds of millions of new users – highlights the stakes involved. India's digital payments success story becomes a liability if the infrastructure enabling it proves vulnerable to AI-powered attacks that can analyze, exploit, and propagate across systems faster than human defenders can respond.

What makes AI-powered cybersecurity threats particularly challenging is their dual-use nature. The same capabilities that make Claude Mythos valuable for defensive security research – analyzing complex systems, identifying subtle vulnerabilities, understanding attack chains – make it dangerous in attackers' hands. Unlike traditional security tools that clearly separate offensive and defensive uses, AI models don't inherently align with good or bad actors, only with the capabilities that amplify whoever deploys them.

The path forward requires investment in AI-powered defenses that match potential AI-powered attacks in speed and sophistication, but also recognition that technology alone won't solve problems rooted in system complexity, legacy infrastructure, and the fundamental challenge of securing systems that must remain accessible to hundreds of millions of legitimate users while blocking determined attackers.

Finance Minister Sitharaman's decision to convene this high-level meeting signals appropriate concern. The banking sector's challenge isn't whether AI will transform cybersecurity threats – it already has. The challenge is building defenses that operate at AI speed, detecting and responding to threats measured in milliseconds rather than hours, because when attackers deploy AI, human response times become inadequate regardless of expertise. And, our Minister's concern is justified – if Mythos can find 271 vulnerabilities in one of the most scrutinized open-source codebases in the world, what could it find in banking systems?

–Authored by Govind Rammurthy, CEO & Managing Director, eScan