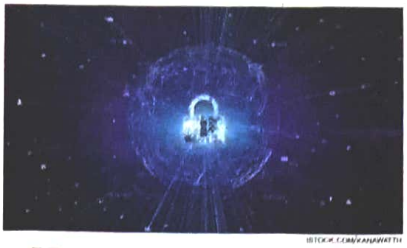

Newest AI model can pick holes in every system

In such a scenario, the wisest companies are those that invest in an AI-led defence strategy, says **Mathures Paul**

US treasury secretary Scott Bessent and US Federal Reserve chair Jerome Powell recently summoned an urgent meeting of Wall Street leaders to express concerns that the latest AI tool — Claude Mythos Preview — from Anthropic will usher in an era of greater cyber risks. The urgency of the meeting highlights a shift from AI-related optimism tech companies portray.

Claude Mythos Preview has found thousands of high-severity vulnerabilities, including some in every major operating system and web browser, said Anthropic. Such is the fear that Anthropic launched the cybersecurity AI model to a select group of players, including Amazon, Apple and Microsoft. The list of vetted organisations includes Broadcom, Cisco, Nvidia and CrowdStrike. The vetted coalition, known as Project Glasswing, also has some of Anthropic's competitors in AI such as Google and organisations that maintain critical open-source software, such as the Linux Foundation.

"Mythos represents the transition from generative AI (text and code generation) to agentic AI (autonomous action). We are moving away from users asking a model to write a script to a world where AI acts as an autonomous agent. It can scan, identify and exploit vulnerabilities without human intervention. We are seeing a 100x speed advantage for attackers. Mythos isn't just a smarter model; it's an automated factory for zero-day exploits. This marks the end of security through obscurity," Neehar Pathare, MD, CEO and CIO of 63SATS Cybertech, told **The Telegraph**. The company, headquartered in Mumbai, offers advanced cybersecurity solutions.

Pathare spoke of "superhuman reasoning". "It isn't just a faster scanner; it's identifying bugs that have remained undetected for decades. This shifts AI interaction from creative assistant to autonomous auditor."

The scale of potential disruption is difficult to imagine. Let's look at an unconnected cybersecurity scenario. In June 2024, several major London hospitals had to cancel operations and blood transfusions after a cyberattack. Over 10,000 appointments were cancelled. Such cyberattacks are rare, but with new AI models this could change, especially if bad actors and state-backed hackers gain access to similar tools.

"In the wrong hands, a model that can find every major browser and OS vulnerability could cripple global supply chains, financial grids and healthcare systems. The AWS report about 600-plus devices compromised by a low-skill hacker using current tools (Claude and DeepSeek) is a terrifying proof of concept for what Mythos can do," said Pathare.

He added that while keeping such powerful tools in the hands of a few technology giants or governments sounds safer, it creates a monopoly defence.

Anthropic, which was founded in 2021 is racing to build increasingly powerful AI systems. The company has been in the news after the Pentagon deemed it a supply-chain risk this year for demanding certain limitations on the use of its technology. A federal judge later stopped the designation from going into effect.

AI & You

Here's an unconnected cybersecurity scenario. In 2024, London hospitals had to cancel surgeries, blood transfusions after a cyberattack

Researchers have not been given access to independently verify Anthropic's claims about Mythos's performance. The safeguards for the AI model are a work in progress, according to Anthropic.

So where do Indian companies stand? Said Pathare, "They are in a transition phase. While the top-tier BFSI and IT sectors are adopting AI-led resilience, many mid-market firms are still relying on last decade's playbook, firewalls and manual patching, which are useless against an agentic attacker." He added, "Indian firms face an unenviable dilemma — do they subject their indigenous tech stacks to an audit by a foreign AI (Mythos), or risk remaining vulnerable to those who eventually will?"

It's not all doom. Software makers now have a "God-mode" for debugging. Said Pathare, "The future of cybersecurity will be AI versus AI. The winner will be the one with the most agile AI-led defence strategy."

Sure enough, days after Mythos arrived, OpenAI introduced GPT-5.4-Cyber, which aims to find software issues so organisations can fix them. The model will be offered to some participants of OpenAI's Trusted Access for Cyber programme.