Indian Express,dt:1.5.26,pg.no.12

PADMINI DHRUVARAJ @ Bengaluru

Mythos identifes thousands of so-called zero-day vulnerabilities, flaws that were previously unknown even to software's own developers

MYTHOS

The AI that knows your software's secrets before you do

ON April 7, American AI company Anthropic quietly made an announcement that shook the world's cybersecurity community far more than most headlines about artificial intelligence ever do. It said it had built a new AI model so capable of breaking into computer systems that it had decided not to release it to the public at all. The model is called Claude Mythos.

Mythos is a general-purpose AI model, not one specifically designed for security work. But during testing, Anthropic found that it possesses cybersecurity abilities that far exceed any prior model it has built. To understand that in plain terms, Mythos can thoroughly scan the software running your phone, your bank's servers, your government's databases, and find the hidden cracks and figure out how to break in automatically, without a human guiding it at any step.

Over the past few weeks of testing, Anthropic used Mythos to identify thousands of so-called zero-day vulnerabilities, flaws that were previously unknown even to the software's own developers, in every major operating system and every major web browser. Some of these flaws had been sitting undiscovered for years. One previously unknown vulnerability in OpenBSD, an operating system famous for its security hardening, had been around for 27 years.

What makes this particularly striking is that Mythos was not built to do any of this. Anthropic has said: "We did not explicitly train Mythos Preview to have these capabilities. Rather, they emerged as a downstream consequence of general improvements in code, reasoning, and autonomy." In other words, Anthropic built a very clever general-purpose AI, and it turned out to be extraordinarily good at hacking as a side effect.

What Anthropic did with it

Rather than shelve the model or release it openly, Anthropic chose a third path. It launched Project Glasswing, an initiative to secure the world's most critical software, bringing together Amazon Web Services, Apple, Broadcom, Cisco, CrowdStrike, Google, JPMorgan Chase, the Linux Foundation, Microsoft, NVIDIA, and Palo Alto Networks as launch partners. The idea is to give these companies access to Mythos so they can find and fix vulnerabilities in their own software before bad actors get hold of a similar tool.

Anthropic committed up to $100 million in usage credits for Mythos, as well as $4 million in direct donations to open-source security organisations.

The company has also been clear that it believes this window will not stay open for long. Anthropic estimates that similar capabilities will proliferate from other AI labs within 6-18 months.

Where India stands

No Indian company was included in the list of organisations given access to Mythos under Project Glasswing.

Finance Minister Nirmala Sitharaman confirmed that the matter is being addressed at the highest levels, stating that the Ministry of Electronics and Information Technology is actively engaging with the US administration, Anthropic, and organisations already testing the model.

Industry body Nasscom wrote to Anthropic seeking inclusion of Indian firms in Project Glasswing, highlighting the need for Indian companies to access such tools to strengthen global cybersecurity resilience, especially as their software-supported systems are worldwide.

India's banking system, power grids, and telecom networks all run on software stacks that contain the same kinds of vulnerabilities Mythos is designed to find. The Finance Minister asked the Indian Banks' Association to develop a coordinated mechanism for quickly responding to threats linked to the model and directed banks to strengthen their cybersecurity systems.

Security analysts describe the Mythos announcement as a reminder that AI has come a long way in just a few years, and that the baseline has genuinely shifted. The question is no longer whether AI will change cybersecurity, but whether defenders can move faster than attackers.