Concerns grow over Mythos AI

Anthropic's new Mythos AI model is raising concerns after the company said the preview had found thousands of serious vulnerabilities in every major operating system and web browser



CONCERNS

- Experts warn model can identify and exploit ***previously unknown vulnerabilities*** faster than companies can repair them
- Advanced coding and autonomous capabilities could speed up ***sophisticated cyberattacks***, particularly in sectors such as banking

MYTHOS DEBUT

- **Apr 7, 2026:** Launch of ***Claude Mythos Preview*** – AI tool designed for defensive cybersecurity tasks

Model rolled out through initiative called ***"Project Glasswing"***

- Access granted to tech majors – including ***Amazon***, ***Microsoft***, ***Nvidia*** and ***Apple***

- Access also extended to group of ***more than 40 other organisations*** that build or maintain critical software infrastructure

- During debut, Anthropic said model's ***ability to find software flaws*** could, if misused, pose serious risks to economies, public safety and national security

- ***White House*** and Anthropic CEO ***Dario Amodei*** *(right)* discussed working together amid fears over Mythos



- U.S. government planning to make version of Mythos available to ***major federal agencies***

- Model raised alarm bells in UK, with authorities holding talks with ***banks and cybersecurity officials*** to assess possible risks

Source: Reuters Pictures: Getty Images © GRAPHIC NEWS