ADVISORY: ANTHROPIC CLAUDE MYTHOS PREVIEW STATEMENT

16 APRIL 2026

Artificial Intelligence company, Anthropic, has recently published an article announcing the development of a powerful new general-purpose language model with striking capabilities in the way hardware and software vulnerabilities are identified and patched. Of note, this new model, Claude Mythos Preview, possesses a much higher success rate in identification and exploitation of vulnerabilities than previously seen in other models.

Anthropic has not publicly released the model, instead, creating Project Glasswing which has provided several high-profile companies with access to test Mythos Preview, and identify and patch vulnerabilities in critical infrastructure. Due to this responsible restricted approach, independent evaluation of the model has not yet been undertaken, but it is projected to considerably contribute to changes in the cyber threat landscape.

Although Anthropic has not released this model, organisations should expect vulnerability disclosures to accelerate over the coming months as similar AI language models are developed and released for public consumption.

Whilst there is no immediate threat, the CSC would like to remind business leaders and technical teams to ensure that their systems and services are protected.

We have provided a short list of key security controls for organisations to consider on the following page. We highly recommend business leaders and technical teams to implement the appropriate measures to improve their security posture.

ADVISORY:
ANTHROPIC CLAUDE MYTHOS PREVIEW STATEMENT

Harden your environments, prioritise your patching discipline and assess your exposures. The CSC recommends organisations to apply key security controls:

- **Defence-in-depth** – Reduce your attack surfaces by removing exposed systems and services, apply strict access controls wherever possible and utilise a defence-in depth approach.
- **Patch Management** – Apply patches and security without delay. Review and optimise your patch management processes to increase the pace of response to disclosed vulnerabilities.
- **Understand Your Assets** – Regularly review and maintain inventories of your hardware and software assets. Use this information to effectively respond to vulnerability and exposure issues.
- **Logging & Monitoring** – Perform comprehensive logging and monitoring of devices and systems using modern detection and response tools.
- **Mitigation Preparation** – Ensure you are prepared to perform mitigation measures in a timely fashion to limit the impact of compromise should it occur.
- **Incident & Continuity Plans** – Ensure your cyber incident and business continuity plans are reviewed at regular intervals and exercised in preparation for any incident.
- **Board-Level Engagement** – Effective governance requires cyber risk to be actively owned and prioritised at board-level.

The CSC website has a wide range of cyber security topics in our Advice & Guidance section: https://csc.gov.im/advice-guidance/