Telecoms operators and vendors should prepare for the impact that LLMs will have on cyber security

April 2026

Joseph Attwood

On 7 April 2026, Anthropic announced the launch of its new large language model (LLM), Claude Mythos Preview, which Anthropic claims is unparalleled in its ability to find and exploit software vulnerabilities. The company and the media claim that this represents a major leap forward in cyber-security capabilities, but these assertions appear somewhat exaggerated. Nonetheless, LLMs are gradually becoming increasingly effective at vulnerability detection and exploitation; this will have implications for telecoms operators and vendors.

Claude Mythos Preview likely represents a gradual, rather than a step-change, improvement in LLMs' cyber-security capabilities

Anthropic highlights improved cyber-security capabilities as the standout feature of Claude Mythos Preview. The company claims that its was able to use the LLM to identify thousands of zero-day vulnerabilities in the software it analysed.[1] Reportedly fearful of the implications of malicious actors using this LLM, Anthropic has not made it available to the public. Instead, the company formed Project Glasswing in partnership with organisations such as AWS, Broadcom, Google, JPMorgan Chase, as well as others. Project Glasswing gives its partners access to Claude Mythos Preview in the hope that they can use it to find and fix vulnerabilities in their software before malicious actors get the chance to exploit them.

Is Anthropic's claim that the model cannot be released to the public due to the harm that malicious actors could do using the model underpinned by legitimate concerns, or is this just an unfounded fear or a marketing stunt to hype up the model's capabilities? Good and bad actors have already been making use of the cyber-security capabilities of LLMs for some time. Testing by the UK's AI Security Institute (AISI) found that Claude Mythos Preview scored as the leading or one of the leading models in terms of its security capabilities compared to other LLMs.[2] However, while the model scored highly, the improvement in capabilities compared to other existing models was not especially dramatic. Instead, Claude Mythos Preview's capabilities seem to roughly track with the historic rate of improvement in the security capabilities of frontier LLMs. Additionally, other researchers have found that other LLMs – including much smaller ones – can

[1] Anthropic did not state the amount of software that was analysed.

[2] AISI (2026), *AISI's evaluation of Claude Mythos' capabilities*.

also identify the software vulnerabilities that Anthropic discovered (at least the ones it has released details on).

That being said, the overall direction seems to be clear: LLMs are getting better at detecting and exploiting vulnerabilities.

Operators and vendors should use LLMs to check for security vulnerabilities; however, this will require a nuanced process

Using LLMs to check for vulnerabilities in software, systems and networks requires more than just pointing an LLM at a software repository. Anthropic's testing of Claude Mythos Preview involved launching multiple agents in parallel that had access to the source code and the project running in a container.[3] Each agent analysed code and experimented with the project as it ran, with investigations being prioritised on code files that the model decided were most likely to have a vulnerability. Claude Mythos Preview was then used to confirm that the bugs identified by each agent were real and of high severity. For some pieces of software, Anthropic executed around 1000 agentic workflows in the process of scanning for vulnerabilities.

Using LLMs for vulnerability scanning in an ad hoc, manually directed fashion has the potential to be slow and somewhat ineffectual. Operators and vendors will benefit from implementing tooling and workflows that direct vulnerability scanning to be done in an effective way (for example, by creating agents that each look at different parts of the code) and that automate the process end to end. For example, the tooling would need to automate the launch of containers running the program under test, orchestrate multiple AI agents that each execute focused testing, and use an LLM to validate and consolidate the outputs of each agent.

Red teaming with LLMs can be used by organisations that are writing code.[4] In the telecoms context, this will mainly include vendors as well as advanced operators that do a lot of their own software development. LLMs may also be used to ensure that the open-source software ingested by organisations is secure – this approach cannot be used on proprietary software organisations procure from vendors as it depends on access to the source code. However, it is likely that it will generally be assumed open-source communities (and vendors) have already done this testing on their software. Due to time and cost considerations, only a small proportion of organisations are likely to implement LLM-driven vulnerability scanning into their CI/CD/CT pipelines.[5]

Down the line, operators may be able to use LLMs to detect vulnerabilities in their configuration and deployment artefacts (and potentially their APIs) that they create and/or modify by

[3] Anthropic (2026), *Anthropic's blog on Claude Mythos' capabilities*.

[4] Red teaming is the process by which security researchers attempt to attack a system to uncover vulnerabilities.

[5] Continuous integration, continuous delivery/deployment and continuous testing (CI/CD/CT).

themselves. However, this extends to the very edges of the cyber-security uses cases for LLMs that are currently being explored.

Operators and vendors need to prepare for a world in which the discovery of vulnerabilities happens much more frequently

If LLMs are indeed much better at finding security vulnerabilities than human experts, the rate at which new vulnerabilities are discovered going forward will exceed historical levels.[6] As a result, the rate at which organisations need to ingest security patches from their internal development teams, their vendors and open-source projects will increase. Therefore, it will become increasingly important that operators and vendors have highly automated CI/CD/CT toolchains.

The improved cyber-security capabilities of LLMs will increase the rate at which both good and bad actors can identify security vulnerabilities, meaning this is both an opportunity and a threat for operators and vendors. It is not clear how this will shift the balance between offence and defence, but if organisations do not adapt, they will fall behind AI-assisted malicious actors.

Operators and vendors should begin exploring the use of LLMs in red teaming exercises as a lot of learnings still need to be made, for example, relating to how best to apply LLMs to secure software for networks functions versus BSS/OSS software. The telecoms industry should seek engagement in initiatives like Project Glasswing – or consider forming its own initiatives – to secure the telecoms software supply chain and ensure that best practices for securing this supply chain with LLMs are developed. Telecoms operators and vendors should also consider whether the creation of telecoms-specific LLMs would enable them to achieve better security outcomes.

Network cyber security will be a key topic in Analysys Mason's Network Automation and Orchestration research programme in the coming year. Areas of focus will include using AI for security, securing AI systems and building quantum-safe networks.

[6] This will be exacerbated by organisations using vulnerability-prone AI-generated code.