УДК 004.582:004.946

Teterev M. Y.
Chopliak V. V.

CLAUDE MYTHOS: ANALYSIS OF THE CAPABILITIES OF AUTONOMOUS AI AGENTS IN DETECTING CRITICAL 0-DAY VULNERABILITIES

Vinnytsia National Technical University

Анотація

У роботі досліджено ефективність автономних AI-агентів (на прикладі Claude Mythos) у кібербезпеці. Порівняно підходи AI-аналізу з класичним фаззингом у виявленні 0-day вразливостей. На реальних кейсах (багаторічні критичні баги в OpenBSD та FFmpeg) доведено перевагу автономного пошуку коду. Також окреслено ключові етичні та безпекові ризики використання AI-систем в offensive security.

Ключові слова: AI-агенти, кібербезпека, 0-day, Claude Mythos, фаззинг, автономний аналіз.

Abstract

This paper explores the efficiency of autonomous AI agents (using Claude Mythos as a case study) in cybersecurity. It compares AI-driven analysis with traditional fuzzing for zero-day vulnerability detection. The technical superiority of autonomous search is demonstrated through real-world cases involving legacy bugs in OpenBSD and FFmpeg. Key ethical and security risks of deploying AI in offensive security are also outlined.

Keywords: AI agents, cybersecurity, zero-day, Claude Mythos, fuzzing, autonomous analysis.

Introduction

Artificial intelligence technologies are rapidly transforming modern cybersecurity approaches. While previous language models primarily acted as assistants for developers and security researchers, frontier AI systems are increasingly capable of independently performing complex analytical tasks. One of the most notable examples is Claude Mythos Preview, introduced by Anthropic in April 2026 as part of Project Glasswing.

Research results

One of the main technological differences between Claude Mythos and classical vulnerability detection systems is its autonomous reasoning capability. Traditional fuzzing systems rely on generating large numbers of malformed inputs to trigger crashes or abnormal behavior. In contrast, Mythos analyzes large codebases contextually, searching for logical inconsistencies and hidden exploitation chains.

A significant confirmation of these capabilities was the discovery of a 27-year-old vulnerability in OpenBSD related to TCP selective acknowledgments (SACK). Another notable case involved FFmpeg, where Mythos identified a 16-year-old vulnerability in the H.264 decoder.

Analysis of the results

The emergence of Claude Mythos demonstrates a major shift in cybersecurity methodology. Classical vulnerability detection tools are limited by predefined signatures, heuristics, or brute-force fuzzing approaches. Autonomous AI agents, however, can reason about software behavior and generate new hypotheses regarding possible attack vectors.

At the same time, these capabilities create serious ethical and security concerns. Anthropic deliberately restricted public access to Mythos because the same mechanisms that can protect infrastructure may also be used for offensive cyber operations.

Conclusion

Claude Mythos Preview represents one of the first examples of an autonomous AI agent capable of independently detecting and analyzing critical zero-day vulnerabilities in complex software ecosystems. The discovery of decades-old vulnerabilities in OpenBSD and FFmpeg confirms the practical effectiveness of this approach and demonstrates the growing role of AI in offensive and defensive cybersecurity.

REFERENCES

1. Anthropic Security Research: Claude Mythos Preview. URL: https://red.anthropic.com/2026/mythos-preview/.
2. TechCrunch: Anthropic debuts preview of Mythos. URL: https://techcrunch.com/2026/04/07/anthropic-mythos-ai-model-preview-security/.
3. Tom's Hardware: Mythos identifies thousands of zero-days. URL: https://www.tomshardware.com/tech-industry/artificial-intelligence/anthropics-latest-ai-model-identifies-thousands-of-zero-day-vulnerabilities-in-every-major-operating-system-and-every-major-web-browser-claude-mythos-preview-sparks-race-to-fix-critical-bugs-some-unpatched-for-decades.
4. Nykyporets S. S. Hadaichuk, N. M. Foreign language media literacy as a protective factor against AI-generated disinformation and psychological stress in technical higher education in Ukraine. *Transformational vectors of public administration, law, and humanities in the development of the modern educational system: Scientific monograph.* Riga, Latvia: "Baltija Publishing", 2025. 492 p. P. 305-330. DOI: https://doi.org/10.30525/978-9934-26-647-8-14.
5. Kot S. O., Nykyporets, S. S. Activating students' cognitive engagement in technical English learning with AI tools. *Science and education in the third millennium: information technology, education, law, psychology, social security and work, management. International collective monograph.* Volume I. Institute of Public Administration Affairs. Lublin, Polska, 2025. 532 p., Pp. 295-332. DOI: https://doi.org/10.5281/zenodo.16942267.
6. Ibrahimova L. V., Nykyporets, S. S. Information security in the global context: linguistic perspectives and the role of English. *International security studios: managerial, technical, legal, environmental, informative and psychological aspects.* International collective monograph. Volume II. ISAP, Research and Education. 2025. 436 p., P. 321-345. DOI: https://doi.org/10.5281/zenodo.15356365.
7. Nykyporets S. S. The role of AI in enhancing technical vocabulary acquisition among engineering students. *Distance Education in Ukraine: Innovative, Normative-Legal, Pedagogical Aspects*. 2025. №5. Pp. 79-85. DOI: https://doi.org/10.18372/2786-5495.1.20579.

Тетерев Максим Юрійович – студент групи 5ПІ-25б, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, м. Вінниця, e-mail: teterevm08@gmail.com.

Науковий керівник: ***Чопляк Вікторія Володимирівна*** – викладач англійської мови, кафедра іноземних мов, Вінницький національний технічний університет, м. Вінниця, e-mail: nikavnuchkova@gmail.com.

Maksym Y. Teterev – Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: teterevm08@gmail.com.

Scientific Supervisor: ***Victoria V. Chopliak*** – teacher of English, Foreign Languages Department, Vinnytsia National Technical University, Vinnytsia, e-mail: nikavnuchkova@gmail.com.