TO: XXXXXXXX XXXXXXX
XXXXXXXX@openai.com
Trustworthy AI Team
OpenAI

FROM: Andrew Lawrence
contact@exton.info
Tufts University
Tufts AI Safety Student Association

RE: “Slopsquatting,” an Emerging Supply Chain Cyber-Attack Taking Advantage of ChatGPT Hallucinations in Code Suggestions

---

Executive Summary
AI models like ChatGPT are now central to large-scale software development. This widespread adoption introduces a novel and under-recognized attack surface. One key threat is “slopsquatting,” where these Large Language Models (LLMs) hallucinate non-existent software package names, which malicious actors subsequently register, leading to security breaches when unsuspecting developers incorporate these fictitious packages into their codebases.

OpenAI’s Stake
OpenAI’s reputation and market dominance depends on the reliability and security of its frontier models. Slopsquatting undermines that trust by introducing security vulnerabilities that can lead to compromised systems. Not only is this damaging for OpenAI’s enterprise clientele, but associations with security lapses erode general user confidence and deter adoption.

Recommendations
Even with lower hallucination rates than competitors, OpenAI must act to address this risk:

1. **Implement in-model verification processes:** to confirm package existence in real-time against official registries during code output.
2. **Introduce deployment-stage slopsquatting assessments:** to evaluate and monitor model outputs for hallucinated dependencies before release.
3. **Initiate collaboration with Package Registries:** to share intelligence and preempt malicious uploads, enhancing detection and prevention of malicious code execution.

Implications
Should OpenAI want to maintain its edge over competitors, then they must lead a campaign against Model Injection Attacks. Doing so will reaffirm their commitment to trustworthy AI systems and demonstrate the seriousness of ChatGPT as an enterprise technology.

Conclusion
The proposed measures offer a balanced approach. By proactively implementing these policies, OpenAI can interlock policy safeguards, mitigating risk and upholding its commitment to trustworthy AI. Not only would OpenAI address a current vulnerability by doing so, but also set an industry precedent for responsive cyber-security practices.

Sources

Castañeda, Ari. "UTSA Researchers Investigate AI Threats in Software Development." Utsa.edu, 7 Apr. 2025, www.utsa.edu/today/2025/04/story/utsa-researchers-investigate-AI-threats.html.

Claburn, Thomas. "AI Code Helpers Just Can't Stop Inventing Package Names." Theregister.com, The Register, 30 Sept. 2024, www.theregister.com/2024/09/30/ai_code_helpers_invent_packages/.

Claburn, Thomas. "AI Hallucinates Software Packages and Devs Download Them – Even If Potentially Poisoned with Malware." Theregister.com, The Register, 28 Mar. 2024, www.theregister.com/2024/03/28/ai_bots_hallucinate_software_packages/.

Spracklen, Joseph, et al. "We Have a Package for You! A Comprehensive Analysis of Package Hallucinations by Code Generating LLMs." ArXiv.org, 2024, arxiv.org/abs/2406.10279.

Toulas, Bill. "AI-Hallucinated Code Dependencies Become New Supply Chain Risk." BleepingComputer, 12 Apr. 2025, www.bleepingcomputer.com/news/security/ai-hallucinated-code-dependencies-become-new-supply-chain-risk/.