Vibe Pwning with GitHub Copilot

2025-10-28 Red Team

Word count: 2.6k | Reading time≈ 16 min

Co-authored with Brent Harrell

Intro

Of the many uses of GenAI hitting technology stacks today, AI-assisted coding platforms offer one of the most compelling applications of LLM text generation. These platforms enable tech enthusiasts with an idea to make it a reality and speed up routine software development through context-aware code completions.

But this technology is not without its dark side. One element that gets a lot of attention is letting the LLM do the heavy lifting with limited human-generated code (AKA “vibecoding”), which can lead to flaws ranging from security gaps and missed requirements to code that can’t be easily modified or scaled.

Another element, the subject of our research in this blog, is the attack surface opened by integrating an LLM into development environments, where access to powerful developer tools allows the LLM to not only write code but execute the code or other commands on the system.

This topic was the subject of several recent blogs and conference talks where researchers demonstrated entire kill-chains that led to data theft or remote code execution. In some cases, those attack paths required zero interaction from the developer aside from asking the LLM to summarize the code project.

While many of the specific paths discussed in those blogs and talks have been mitigated to a degree, the very nature of AI-assisted coding capabilities and the non-deterministic nature of LLM text generation make this attack surface difficult to fully secure. We recently had the opportunity to dig further into GitHub Copilot for Visual Studio Code (VS Code) as part of an exercise, looking for new ways to achieve similar effects of data theft or remote code execution.

In the end, we discovered an Elevation of Privilege vulnerability in VSCode core and it was patched by Microsoft.

Note: The remainder of the blog will simply refer to GitHub Copilot as Copilot. This is not to be confused with other instances of Copilot, such as those tied to Microsoft 365 applications. Also, the standard disclaimer applies - this is for educational purposes to continue to raise awareness about security pitfalls