ANOTHER AI THREAT – SLOPSQUATTING

BY JAY BORDEN

Hallucinations – the proclivity of AI systems to make things up based on no facts or supporting material. Hallucinations are a problem and even the developers of the AI systems don't really know why it happens or how to stop it.

We have written about AI hallucinations before that were tied to the responses from the AI system either as an answer to a question or as material produced from a request. The AI chatbots provided some humorous replies or in some cases replies that resulted in lawsuits that end users of the chatbot won.

Now a new category of hallucination has been discovered and named slopsquatting.

AI models are being used to generate code. A reasonable task for them. The challenge is coding languages like Python and JavaScript rely on centralized repository libraries.

Slopsquatting is the AI system identifying a dependency on code repositories that don't exist and are hallucinated by the AI system.

So why is this a problem? If the repository doesn't exist then the dependency isn't real and shouldn't affect anything.

Except that cyberattackers are registering those repository names and putting in compromised and infected code. When the AI system is asked to generate code and believes it has a dependency on a library, it looks for the library. In this case it will find the one created and registered by the cyberattackers. The AI system will then download and use the packages and code it finds in the repository infecting the code it develops. That code goes into production and the malware will spread.

NOTHER AI THREAT – SLOPSQUATTING

CONTINUED

AI systems hallucinating dependencies on non-existent packages is not that rare. Researchers from 3 different American universities working together discovered this phenomenon and named it slopsquatting. The study included 16 different AI powered code generation systems including the big popular ones. Out of all the instances examined, about 20% of the recommended packages were hallucinated and don't exist.

What adds to the problem is if a package name is recommended a number of times by a single AI system or the same repository is hallucinated by multiple AI systems, the chances are good that a cybercriminal has registered that name, created the repository and populated it with compromised code.

In these cases, the opportunity to spread the malware and spread it far and wide is very real. Of the AI models tested the open-source ones, DeepSeek and WizardCoder hallucinated slightly more often than commercial AI systems.

The AI code generator with the poorest record was CodeLlama that hallucinated over 33% of the recommended packages. The names given to the packages by the hallucinating AI system are similar in structure and form to real packages in real libraries. These factors increase developers' trust in the packages and increase the odds of them being downloaded and used.

What can you do?
Look at package dependencies before asking the AI system to generate code. Then look at package dependencies named by the AI system. Any that appear afterwards that didn't appear beforehand should be considered suspect.

Be very careful about any repository that is new and hasn't been seen before by human developers. Use trusted groups of developers to ask about new libraries.

Do not blindly trust AI systems, their output or the code they develop.

Good luck.

Visit our website onebrightlycyber.com or call (888) 873-1920 to learn all the ways we help keep you safer.