AI DATA POISONING RISKS

BY JAY BORDEN

Let's start with a definition of data poisoning. It is corrupting the data used to train the AI system that it then uses to create it's results. Let's consider a simple example. All sales data is entered in your systems. AI is used to find the most successful sales reps and target accounts. Sounds helpful. But what if someone breaks into your data and changes things? Then the results will be wrong, and you may be promoting the wrong person, targeting the wrong accounts, etc.

Another example. Retailers use AI systems with Chatbots to interact with customers on the website. What if attackers change the AI system so that every time someone inputs the word "problem" the system replies with "We don't care. It's your problem." It may help your competitors but not your reputation.

One more. AI is used in the medical industry to analyze test results and imaging to assist doctors in making diagnoses. What if the AI system is hacked and makes wrong diagnoses? Or ignores certain critical factors and fails to report on a life threatening situation?

On the cybersecurity front one use of AI is to identify phishing emails. What if it is changed to allow those through? Employees would assume that if the email is received the filtering system decided it was safe, and they clicked on the links?

Who will be bear responsibility in these situations? Certainly not the attackers.

Remember that while AI stands for Artificial Intelligence, it has little real intelligence. It doesn't know what it is doing.

AI systems work from what has been termed training data. This is the initial data given to the AI system as input and what it uses in its analysis or determining output.

AI DATA POISONING RISKS

CONTINUED

Newer AI systems continue to learn as they see what other information they can find. Both of these situations provide opportunities for attackers to corrupt the information.

Everyone has heard to not believe everything you read on the internet. AI systems do not have that level of discernment. An AI system will use what it finds to “improve” what it delivers back as answers or output.

The intent of using AI is to reduce human effort and get faster and better results than what people can do. Asking people to verify the findings of the AI system is counter-productive. This is what makes poisoned data so dangerous. Unless the results are blatantly wrong, they will be accepted.

As AI systems are used by so many companies, poisoning the data will impact all who ask questions or rely upon the AI system where that corrupted data influences the AI reply. In this way poisoning the AI system can be viewed as a supply chain breach. No need to breach many companies, just the AI system they reply upon.

The integrity of the AI training data initially and subsequently is essential to trusted results. The more critical the application of the AI system the higher the confidence in the data must be. In my first example above, having a chatbot give rude answers is certainly important to the company, but it’s not as important as getting bad medical diagnoses due to corrupt data.

Companies are already using AI systems. The number of companies using it and the number of applications with AI in each company will only increase. As dependence on the AI system increases, so must trust in its results.

AI poisoning will be exploited by cyber attackers. Companies need to consider this a significant threat and begin determining how to prepare and protect their data and their company.

Knowing risks helps you stay safer.

Visit our site, onebrightlycyber.com to learn all the ways we help keep you safer.

Welcome to peace of mind.

onebrightlycyber.com
(888) 773-1920