Artificial intelligence may put private data at risk

August 3 2018, by Melanie Lefkowitz

Machine learning – a form of artificial intelligence in which computers use data to learn on their own – is rapidly growing and poised to transform the world. But current models are vulnerable to privacy leaks and other malicious attacks, Cornell Tech researchers have found.

Used for everything from predicting what customers want to buy to identifying people at risk for a certain disease, machine learning models are "trained," or taught to perform specific tasks, by processing large sets of data.

Vitaly Shmatikov, professor of computer science at Cornell Tech, developed models that determined with more than 90 percent accuracy whether a certain piece of information was used to train a machine learning system. This could potentially expose sensitive genetic or medical information, detailed data about people's habits or whereabouts, and more.

"If I can figure out if a patient's record was used for a health care study associated with a particular disease, then I can figure out whether that person has the disease," said Shmatikov, whose paper, "Membership Inference in Machine Learning," received the Caspar Bowden Award for Outstanding Research in Privacy Enhancing Technologies, awarded at the Privacy Enhancing Technologies Symposium in July. "This information is very sensitive, and it makes people very nervous if you can discover that their information was used."

Tools that allow you to figure out if a record was used to train an algorithm can be helpful, he said, for those trying to figure out if their data was misused, such as when information from Facebook was acquired by Cambridge Analytica.

In the paper, co-authored with Reza Shokri and Marco Stronati, then Cornell Tech postdoctoral researchers, and computer science doctoral student Congzheng Song, the researchers focused on cloud services from Google and Amazon, which help customers build machine learning models from their own data. Google and Amazon don't reveal how these machine learning tools work, but Shmatikov and his team constructed "shadow models" built from real or fake data that identified the records used to construct them with high accuracy, showing that customers who use these services can easily end up revealing their own training data.

Among the reasons these systems are vulnerable, Shmatikov said, is that the machines may be learning more than intended. In their 2017 paper, "Machine Learning Models That Remember Too Much," Song, Thomas Ristenpart, Cornell Tech associate professor of computer science, and Shmatikov examined how a change to training data before it's processed could cause a machine learning model to memorize and potentially leak the information.

The people creating machine learning models generally consider only whether they work, and not whether the computer is learning more than it needs to know, Shmatikov said. For example, a program that uses images of people to learn to identify a certain visual characteristic, such as eyeglasses, may also be memorizing entire faces.

"We can tell whether a machine learning model has learned how to perform its task, but today we really have no way of measuring what else it has learned," he said. "Our hope is when people are developing machine learning technologies they don't just focus on the basic question

of, 'Does this do what I want it to do?' but they also ask, 'Does it leak information, is it vulnerable to integrity attacks, is it vulnerable to being subverted by participants in malicious ways?' I think this will result in much more robust and interesting machine learning models, and I think this is starting to happen."

Other projects his team is pursuing include privacy risks in collaborative machine learning systems – those that are built jointly by multiple participants – and vulnerabilities in federated learning, where machine learning models are crowdsourced by as many as millions of users.

"Pretty soon, all apps and services that use raw data are going to be using machine learning," he said. "We're trying to better understand how privacy is going to evolve when machine learning becomes ubiquitous."

More information: Machine Learning Models that Remember Too Much: www.cs.cornell.edu/~shmat/shmat_ccs17.pdf

Membership Inference Attacks Against Machine Learning Models: www.cs.cornell.edu/~shmat/shmat_oak17.pdf

Provided by Cornell University