Artificial intelligence cyber attacks are coming – but what does that mean?

August 28 2017, by Jeremy Straub

Credit: AI-generated image (disclaimer)

The next major cyberattack could involve artificial intelligence systems. It could even happen soon: At a recent cybersecurity conference, 62 industry professionals, out of the 100 questioned, said they thought the first AI-enhanced cyberattack could come in the next 12 months.

This doesn't mean robots will be marching down Main Street. Rather, artificial intelligence will make existing cyberattack efforts – things like identity theft, denial-of-service attacks and password cracking – more powerful and more efficient. This is dangerous enough – this type of hacking can steal money, cause emotional harm and even injure or kill people. Larger attacks can cut power to hundreds of thousands of people, shut down hospitals and even affect national security.

As a scholar who has studied AI decision-making, I can tell you that interpreting human actions is still difficult for AI's and that humans don't really trust AI systems to make major decisions. So, unlike in the movies, the capabilities AI could bring to cyberattacks – and cyberdefense – are not likely to immediately involve computers choosing targets and attacking them on their own. People will still have to create attack AI systems, and launch them at particular targets. But nevertheless, adding AI to today's cybercrime and cybersecurity world will escalate what is already a rapidly changing arms race between attackers and defenders.

Faster attacks

Beyond computers' lack of need for food and sleep – needs that limit human hackers' efforts, even when they work in teams – automation can make complex attacks much faster and more effective.

To date, the effects of automation have been limited. Very rudimentary AI-like capabilities have for decades given virus programs the ability to self-replicate, spreading from computer to computer without specific human instructions. In addition, programmers have used their skills to automate different elements of hacking efforts. Distributed attacks, for example, involve triggering a remote program on several computers or devices to overwhelm servers. The attack that shut down large sections of the internet in October 2016 used this type of approach. In some

cases, common attacks are made available as a script that allows an unsophisticated user to choose a target and launch an attack against it.

AI, however, could help human cybercriminals customize attacks. Spearphishing attacks, for instance, require attackers to have personal information about prospective targets, details like where they bank or what medical insurance company they use. AI systems can help gather, organize and process large databases to connect identifying information, making this type of attack easier and faster to carry out. That reduced workload may drive thieves to launch lots of smaller attacks that go unnoticed for a long period of time – if detected at all – due to their more limited impact.

AI systems could even be used to pull information together from multiple sources to identify people who would be particularly vulnerable to attack. Someone who is hospitalized or in a nursing home, for example, might not notice money missing out of their account until long after the thief has gotten away.

Improved adaptation

AI-enabled attackers will also be much faster to react when they encounter resistance, or when cybersecurity experts fix weaknesses that had previously allowed entry by unauthorized users. The AI may be able to exploit another vulnerability, or start scanning for new ways into the system – without waiting for human instructions.

This could mean that human responders and defenders find themselves unable to keep up with the speed of incoming attacks. It may result in a programming and technological arms race, with defenders developing AI assistants to identify and protect against attacks – or perhaps even AI's with retaliatory attack capabilities.

Avoiding the dangers

Operating autonomously could lead AI systems to attack a system it shouldn't, or cause unexpected damage. For example, software started by an attacker intending only to steal money might decide to target a hospital computer in a way that causes human injury or death. The potential for unmanned aerial vehicles to operate autonomously has raised similar questions of the need for humans to make the decisions about targets.

The consequences and implications are significant, but most people won't notice a big change when the first AI attack is unleashed. For most of those affected, the outcome will be the same as human-triggered attacks. But as we continue to fill our homes, factories, offices and roads with internet-connected robotic systems, the potential effects of an attack by artificial intelligence only grows.

This article was originally published on The Conversation. Read the original article.

Provided by The Conversation

Citation: Artificial intelligence cyber attacks are coming – but what does that mean? (2017, August 28) retrieved 12 May 2024 from https://phys.org/news/2017-08-artificial-intelligence-cyber.html