Testimony before the House Homeland Security Subcommittee on Cybersecurity, Infrastructure Protection, and Innovation

Securing the Future: Harnessing the Potential of Emerging Technologies While Mitigating Security Risks

June 22, 2022

Dr. Andrew Lohn

Chairwoman Clarke, Ranking Member Garbarino, and members of the Subcommittee, thank you for the opportunity to testify today. I am Andrew Lohn, Senior Fellow in the CyberAI Project of the Center for Security and Emerging Technology at Georgetown University. It is an honor to be here. During the next few minutes, I would like to discuss a few of the ways that artificial intelligence intersects with cybersecurity.

To start, it is worth being clear about what makes these two related topics different. Cybersecurity is about protecting the digital world from miscreants, and AI is just one part of that digital world. What distinguishes AI capabilities from more traditional technology is when they perform tasks that until recently required a human—such as a "smart" refrigerator that sees what's on its shelves and suggests a recipe, or an AI-assisted computer that helps drive a car.

AI for Cybersecurity

The distinction between cyber and AI does get murky sometimes. For one, some of the most promising AI systems can help protect digital systems. That has been true for many years in the fight to detect spam or phishing emails – and the capabilities continue to improve to keep pace with attackers.

Another area where AI has shown promise is in detecting attackers once they're in the network, which is known as intrusion detection. Hackers often try to act like normal users and write their malware to blend in with normal software, but there are usually subtle differences that AI can detect to weed them out. This too requires a continual stream of new advances to keep up with attackers who are constantly adapting.

AI Needs Cybersecurity

At the same time, AI systems are digital too, so they need their own cybersecurity protections. While AI-enabled systems have similar vulnerabilities to other types of software, they also have their own unique vulnerabilities. They learn to recognize patterns in data, such as which aspects of an image represent a dog, or which streams of data between two computers are benign and which are malicious. But a clever attacker can change the image or the data stream to fool the AI. There are also ways to trick the AI into revealing data that is meant to remain private. Further, the systems are vulnerable throughout the design process. AI is usually assembled from publicly available components like data, programming libraries, and other AI models that can all potentially be compromised.

AI Subverts Cybersecurity

While AI needs cybersecurity protections, it can also be a means to create new cybersecurity problems. In rare cases, AI might be used to create disruptions in the digital world such as by finding security holes or by helping disguise a digital intrusion. But I'd like to highlight how AI threatens to move beyond the digital world to disrupt our society. AI is able to create images and videos of fake people, or of real people doing or saying things they never said or did. These deepfakes receive a lot of attention, deservedly so, but AI's ability to write text is equally concerning and gets less attention.

Several of the most powerful AI systems today are dedicated to writing text, and they are convincing enough to shift people's stance on important national security topics. CSET's report "Truth, Lies, and Automation" illustrated this point: We used one such system to write tweet-length messages that either supported or opposed sanctions on China, and that either supported or opposed withdrawal from Afghanistan. In a controlled environment, we then showed volunteers a sample of five messages each and measured whether it shifted their opinions.

Comparing the group that read pro-withdrawal messages to the group that saw anti-withdrawal messages, they were 50 percent more likely to want to remove troops and 30 percent less likely to want to maintain troop levels. The Chinese sanctions topic was even more dramatic. In the control group that didn't read any messages, just over half favored sanctions. After reading the five messages though, that flipped. Almost half the population came to oppose sanctions, twice as many as in the control group.

Although we do not know how long-lasting the effect might be, this technique likely appeals to foreign powers who might want to shape our views and control our collective actions. When we did this study last year, these text generators were carefully-guarded proprietary technologies, but now comparable systems are freely available. They are likely within reach of all dedicated nations and even many technologically sophisticated individuals.

Conclusion

In conclusion, AI systems come with risks but can also pave the way for economic and scientific breakthroughs. Access to these tools should be supported, perhaps through initiatives like the National AI Research Resource, but we should also monitor which countries are acquiring them and for which purposes. We should try to harden our population against future malicious uses by promoting trustworthy sources and media literacy while discouraging the spread of disinformation. At the same time, we need to be careful not to deflate the value of all information. Pairing these societal-level defenses with efforts to understand the vulnerabilities of AI systems and the ways AI can boost cybersecurity will go a long way towards securing the nation.