Testimony before the House Science Subcommittee on Investigations and Oversight and Subcommittee on Research and Technology
Securing the Digital Commons: Open-Source Software Cybersecurity
May 11, 2022
Dr. Andrew Lohn

Chairman Foster, Chairwoman Stevens, Ranking Member Obernolte, Ranking Member Feenstra, and members of the Subcommittees, thank you for the opportunity to testify before you 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 risks related to the artificial intelligence supply chain.

A Culture of Sharing

The AI community has been particularly open to sharing. For example, it cost $500,000 and two and a half years to build the famous ImageNet dataset, but the professor who built it released it to everyone. Then Google and Facebook both released their powerful AI engines. Now thousands of the most powerful AI models are a quick download away. It is truly incredible given that these models often range from thousands to millions of dollars to build – and that's in computing cost alone, without even considering the expertise to design them.

The AI Supply Chain

These datasets, models, and AI programming resources are the building blocks of today's AI systems. In the same way that few bakers today grow their own grain and raise their own hens, most AI developers simply combine ready-made components and tweak them for their new applications. Sometimes the whole process only needs a few lines of code and surprisingly little expertise. This approach allowed Google Translate to improve performance in 2016 while trimming from 500,000 lines of code down to just 500.

Sharing has driven both scientific and economic progress, but it has also created an alluring target for attackers.

Supply Chain Vulnerability

For one, an attacker can subvert an AI system by altering the data. That could happen, for instance, by a nefarious online worker while they label the datasets or by a hacker who sneaks into the victim's networks. Alternatively, if the attacker provides a fully trained model, then it can be very hard to find their manipulations.

There is no good way to know if a downloaded model has a backdoor, and it turns out that those backdoors can survive even after the system has been adapted for a new task. A poisoned computer vision system might mistakenly identify objects, or a poisoned language model might not detect terrorist messages or disinformation campaigns that use the attacker's secret code-words.

The programming resources for building AI systems are also vulnerable. Such systems can have thousands of contributors from around the globe writing millions of lines of code. Some of that code has been exploitable in the past. And some of it prioritizes speed or efficiency over security. For example, vision systems need images at a specific size, but the code to resize images allows attackers to swap out one image for another.

And lastly, these resources are only as secure as the organization or system that provides them. Today, the vast majority are hosted in the United States or its allies, but China is making a push to create state-of-the-art resources and the network infrastructure to provide them. If adversaries make the most capable models – or if they simply host them for download – then developers in the United States would face an unwelcome choice between capability and security.

Recommendations

There are a few things Congress can do now to help maximize the benefits of this sharing culture while limiting the security risks that come with it. One step is supporting efforts to provide trusted versions of these AI resources, such as through NIST or the National AI Research Resource. Funding is also needed to do the basic hygiene, cleanup, and audits that are important for security, but that attract few volunteers.

Congress should consider requesting that organizations across the U.S. government create a prioritized list of AI systems and the resources used to build them. This list may be easier to create and maintain if these organizations are incentivized to collect a software bill of materials that lists the components in the software that the government buys or builds.

And lastly, many of these AI systems are new, and so are the attacks on them. The government would benefit from augmenting their red and blue teams of defensive hackers and security specialists with AI expertise to help them discover security holes in our most important systems while also thinking of new, creative ways to subvert them before our adversaries do.

Thank you for the opportunity to testify today, and I look forward to your questions.