Cyber Swords or Shields? Modeling AI's Impact on Cybersecurity

Simple mathematical modeling can add rigor to the largely qualitative debates regarding the potential impact of artificial intelligence (AI) on cybersecurity. Practitioners and policymakers should guide investment and research decisions toward the most beneficial technologies by using modeling to clarify foundational assumptions, assess likely scenarios, and refine predictions of future advances. **This report illustrates this approach with examples focused on phishing, vulnerability discovery, and the race between patching and exploitation.** Further technical details are available in a companion report: https://arxiv.org/abs/2207.13825.

AI technologies can combine the scale of automated phishing with the efficacy of tailored spear phishing. However, CSET predicts only a limited increase in intrusions because attackers may limit their scope of attack to avoid being detected. This is especially true if detection techniques and threat intelligence sharing continue to improve.

AI-enabled vulnerability discovery could benefit defenders or attackers, depending on how the technology progresses. If new automated systems find vulnerabilities faster, they are likely to enable software vendors to fix more vulnerabilities before their products go live, thereby favoring defenders. If new systems find vulnerabilities for longer, akin to the way humans search in live products, then automated systems may favor attackers by generating a larger pool of vulnerabilities for them to use and for defenders to remediate.

AI-enabled patch development could be beneficial. However, technologies that accelerate patch adoption are more likely to generate significant benefits. CSET modeling demonstrates that improving patch adoption by a factor of 5 could reduce the peak number of exposed vulnerabilities by about 25 percent and could decrease the number of exposed computers at the one-hundredth-day mark by 400 percent. **Automated exploit development also has the potential to dramatically change the security landscape.** CSET modeling finds that if both exploits and patches were written instantaneously, then 30 percent more systems would be exposed at the one-hundredth-day mark than they would today without those technologies.

For more information:

- Download the report https://cset.georgetown.edu/publication/will-ai-make-cyber-swords-or-shields
- Contact Us: Andrew Lohn (drew.lohn@georgetown.edu); Krystal Jackson (kaj102@georgetown.edu)