CSET CENTER *for* SECURITY *and* EMERGING TECHNOLOGY

# Summary of *Adversarial Machine Learning and Cybersecurity: Risks, Challenges, and Legal Implications*

**Artificial intelligence systems are vulnerable to attack.** A growing amount of research literature demonstrates that AI models can be manipulated and degraded in ways that are difficult to defend against. AI systems and components are also vulnerable to traditional attack vectors, but AI developers often do not prioritize the security of their models or components. **There is significant ambiguity regarding how existing law covers vulnerabilities in, or attacks on, AI systems, as well as uncertainty about how to better defend them.**

In July 2022, the Center for Security and Emerging Technology, in collaboration with the Program on Geopolitics, Technology, and Governance at the Stanford Cyber Policy Center, convened a workshop of experts in adversarial machine learning, cybersecurity, law, and AI policy to discuss these issues. **This paper provides relevant background and summarizes core recommendations in four high-level areas.** Many of the challenges posed by AI vulnerabilities are social, not technical.

## Recommendations:

**Existing cybersecurity processes can be extended to handle AI vulnerabilities**. AI organizations should adopt risk management frameworks and should experiment with better incorporating AI vulnerabilities into standard cybersecurity processes.

**There is a serious need for greater information sharing**. We encourage the creation of new information sharing arrangements and emphasize that security should be embedded in every stage of the AI life cycle.

**U.S. government agencies should offer greater clarity regarding how AI security concerns fit within their jurisdictions**. In addition, the authors advise against attempts to amend anti-hacking laws to specifically address hacks of AI systems.

**Improving security requires more collaboration**. Adversarial machine learning researchers should work with cybersecurity practitioners to identify realistic threat models to guide research. Federal support for AI research should also more heavily emphasize security, including through application-specific resources and test beds.

## For more information:

- Download the report: https://cset.georgetown.edu/publication/adversarial-machine-learning-and-cybersecurity
- Contact: Micah Musser (Micah.Musser@georgetown.edu)