ECE 491F, Spring 2025:
Computer Software AI security

Meeting time: Mon/Wed/Friday 11:30–12:20pm; **Location:** GAR 112
Prerequisites: ECE 160, ECE 345 (or equivalent)

Ever wondered how to safeguard AI systems from cyber-attacks? Curious about the ways hackers exploit AI vulnerabilities? Welcome to AI Security—a deep dive into the fascinating intersection of artificial intelligence and cybersecurity!

(a) Adversarial Example with Noise (b) Backdoor Attack on Stop Sign

Figure 1: Examples of Vulnerabilities on AI models

In this course, you'll learn:

- How to build different AI models with **PyTorch** Framework.
- How Attackers Target AI: Explore adversarial attacks that fool even the smartest systems. From tricking image recognition to manipulating speech systems, discover the methods behind AI vulnerabilities.
- Defense Mechanisms: Learn cutting-edge techniques to defend against attacks on machine learning models. Build robust systems that stand strong against adversarial inputs and data poisoning.
- Ethics and Privacy in AI: With AI permeating our lives, maintaining privacy and ethical integrity is more crucial than ever. Understand the critical policies and regulations shaping the future of AI security.
- Hands-On Projects: Get practical experience designing secure AI models and implementing defenses in real-world scenarios. Put theory into action with exciting projects that prepare you for a career in AI and cybersecurity.

In this 3-credit course, we will explore the hottest topic in the security community – AI security. You will build on your first experience of hacking AI models, in a way of crafting adversarial examples, and poisoning datasets to get backdoors of AI systems, and explore the up-to-date research papers on designing different safeguard mechanisms.

This is a hands-on experience. **There are no exams.** Instead, every class period will involve you building AI models and exploiting the attacks and defense on the AI models. Over the semester, we will develop the skills required for a project where you will program with **PyTorch** and multiple **AI security toolbox**. **This course provides hands-on experience on AI security, which will enrich your skill set for industry jobs: deep learning, cybersecurity, computer vision, and Applied scientist.**

This course serves as an upper-division Technical Elective, and provides a broad base for future upper-division and graduate-level courses in related areas. The only prerequisites are basic programming and machine learning. Please reach out to your undergraduate advisor if you have any questions about this course or how it fits into your degree plan.

Questions? Contact Dr. Guo: guohanqi@hawaii.edu