CSS 576 Cybersecurity and Artificial Intelligence

Course Overview

The intersection of artificial intelligence (AI) and cybersecurity is emerging as an important field to ensure the integrity of economic and critical infrastructure. Industry and government need more people with training in both AI and cybersecurity.

This course examines the interaction of artificial intelligence and cybersecurity. It covers topics such as data acquisition, model training, applications of AI to cybersecurity, the interaction of AI and humans in security, and securing AI-based systems.

Prerequisites

Familiarity with applied machine learning or cybersecurity concepts is helpful, but not required. The course will provide a high-level introduction to both.

Instructor

Brent Lagesse

lagesse@uw.edu

UW1-260G

Learning Objectives

- Demonstrate secure data acquisition and pre-processing.
- Train AI models to perform or augment security tasks.
- Evaluate the computational and classification performance of AI-based security.
- Assess AI-enhanced cybersecurity mechanisms as an attack surface.

Course Materials

- Course reading material will be derived from openly available research papers describing state of the art work at the intersection of cybersecurity and artificial intelligence.
    - R. S. Siva Kumar *et al*., "Adversarial Machine Learning-Industry Perspectives," *2020 IEEE Security and Privacy Workshops (SPW)*, 2020, pp. 69-75, doi: 10.1109/SPW50608.2020.00028.
    - Bertino, Elisa, Murat Kantarcioglu, Cuneyt Gurcan Akcora, Sagar Samtani, Sudip Mittal, and Maanak Gupta. "AI for Security and Security for AI." In *Proceedings of the Eleventh ACM Conference on Data and Application Security and Privacy*, pp. 333-334. 2021.

    - Xiao, Liang, Xiaoyue Wan, Xiaozhen Lu, Yanyong Zhang, and Di Wu. "IoT security techniques based on machine learning: How do IoT devices use AI to enhance security?." *IEEE Signal Processing Magazine* 35, no. 5 (2018): 41-49.
    - Bozic, Josip, and Franz Wotawa. "Planning the attack! or how to use ai in security testing?." In *Iwaise: First international workshop on artificial intelligence in security*, vol. 50. 2017.
    - Papernot, Nicolas, Patrick McDaniel, Arunesh Sinha, and Michael Wellman. "Towards the science of security and privacy in machine learning." *arXiv preprint arXiv:1611.03814* (2016).
    - Kevin Wu and Brent Lagesse. *Detecting Hidden Webcams with Delay-Tolerant Similarity of Simultaneous Observation.* Pervasive and Mobile Computing, 2020.
    - Cody Burkard and Brent Lagesse. *Analysis of Causative attacks against SVMs Learning from Data Streams.* International Workshop on Security and Privacy Analytics, 2017.
    - Additional papers from top venues such as https://aisec.cc/

- All software used in this course will be freely available

Evaluation

In class exercises (10-15) -- 10%

These exercises will be approximately 30 minutes in duration and include both individual and group work. Each one will take a topic that we discuss in class and allow the students to participate in its hands-on application.

Assignments (4) -- 40%

- Assignment 1: Introduction to commonly used tools implementing a simple AI-Enhanced security mechanism using existing datasets. Example tasks include training and testing a spam filter, intrusion detection system, or malware classifier from existing sources.
- Assignment 2: Design and development of software for data collection, analysis, and preprocessing. Example tasks include reviewing terms of agreement and writing a bot to harvest data from publicly available APIs, analysis of data for bias, and cleaning incomplete data.
- Assignment 3: Implementation and assessment of an AI-enhanced cybersecurity mechanism. Examples tasks include utilizing the data from assignment 2 to implement and assess the performance of the security mechanism. Focus on the assessment differentiates this assignment from assignment 1.
- Assignment 4: Security analysis of the AI-Enhanced cybersecurity mechanism as a novel attack surface. Example tasks involve conducting red team tests on the security mechanism developed in assignment 3 focusing on adversarial examples, data poisoning, and model extraction.

Final Project (1) -- 20%

A large project spanning the full lifecycle of artificial intelligence and security.

Final Exam (1) -- 30%

A comprehensive exam

A 95% is a 4.0 GPA and scale down by 0.1 GPA points for every percentage point less than 95%.

Class Policies

University Policies

Religious Accommodations

Washington state law requires that UW develop a policy for accommodation of student absences or significant hardship due to reasons of faith or conscience, or for organized religious activities. The UW's policy, including more information about how to request an accommodation, is available at Religious Accommodations Policy (https://registrar.washington.edu/staffandfaculty/religious-accommodations-policy/). Accommodations must be requested within the first two weeks of this course using the Religious Accommodations Request form (https://registrar.washington.edu/students/religious-accommodations-request/).

Students with Disabilities

Access and Accommodations: Your experience in this class is important to me. If you have already established accommodations with Disability Resources for Students (DRS), please communicate your approved accommodations to me at your earliest convenience so we can discuss your needs in this course.

If you have not yet established services through DRS, but have a temporary health condition or permanent disability that requires accommodations (conditions include but not limited to; mental health, attention-related, learning, vision, hearing, physical or health impacts), you are welcome to contact DRS at 425-352-5307 or drs@uwb.edu. DRS offers resources and coordinates reasonable accommodations for students with disabilities and/or temporary health conditions.

Reasonable accommodations are established through an interactive process between you, your instructor(s), and DRS. It is the policy and practice of the University of Washington to create inclusive and accessible learning environments consistent with federal and state law.

Academic Integrity

The University takes academic integrity very seriously. Behaving with integrity is part of our responsibility to our shared learning community. If you're uncertain about if something is academic misconduct, ask me. I am willing to discuss questions you might have.

Acts of academic misconduct may include but are not limited to:

- Cheating (working collaboratively on quizzes/exams and discussion submissions, sharing answers and previewing quizzes/exams)
- Plagiarism (representing the work of others as your own without giving appropriate credit to the original author(s))
- Unauthorized collaboration (working with each other on assignments)

Concerns about these or other behaviors prohibited by the Student Conduct Code will be referred for investigation and adjudication by (include information for specific campus office).

Students found to have engaged in academic misconduct may receive a zero on the assignment (or other possible outcome).

Student Conduct

The University of Washington Student Conduct Code (WAC 478-121) defines prohibited academic and behavioral conduct and describes how the University holds students accountable as they pursue their academic goals. Allegations of misconduct by students may be referred to the appropriate campus office for investigation and resolution. More information can be found online at https://www.washington.edu/studentconduct/

Other Policies and Resources

See the School of STEM Course Policies https://www.uwb.edu/getattachment/stem/about/stem-policies/classroom-policies-stem-fc-1-12-17.pdf for additional policies and resources.

Tentative Course Schedule

Week	Topics	Notes	Reading
0	Intro, Cybersecurity Basics and Tools		Cybersecurity AI Survey
1	AI Basics and Tools	Assignment #1 due	AI Overview
2	System Design Requirements		
3	Industry Case Studies	Assignment #2 due	Unbiased Look at Dataset Bias

4	Data Collection, Preprocessing		Poisoning Attacks Survey
5	Model Training and Evaluation (supervised)	Assignment #3 due	
6	Model Training and Evaluation (unsupervised)		ART Documentation
7	System Deployment and Maintenance, Humans in the loop	Assignment #4 due	Adversarial Attacks and Defences: A Survey
8	Incident Response and Lessons Learned		Adversarial Machine Learning -- Industry Perspectives
9	Industry Case Studies		
10	Final Exam	Final Project Due	