|  |  |
| --- | --- |
| **Module 4** | November 19, 2020, 7-9pm CT |
| **Topic:** | Ethics, Fairness, Responsibility, and Transparency in Data-Driven Cybersecurity |
| **Description:** | This module will engage students in hands-on programming assignments, case studies, and discussions to expose them to ethical considerations associated with automated cybersecurity decision-making. |
| **Faculty Lead:** | Blase Ur (Lead) |
| **Asynchronous Content:** | * Asynchronous video 1: Data cleaning and handling missing data (~5 minutes) * Asynchronous video 2: Fairness in cybersecurity ML models (~10 minutes) * Asynchronous video 3: Transparent and explainable ML models for cybersecurity (~5 minutes) * Asynchronous video 4: Defining and actualizing privacy (~10 minutes) * Asynchronous video 5: Collecting data from human subjects (~5 minutes) * Asynchronous video 6: The pitfalls of anonymization (~5 minutes) * Asynchronous video 7: Responsible data lifecycles (~5 minutes) |
| **Synchronous Content:** | * Discussion of privacy definitions and how to actualize privacy for cybersecurity applications in industry * Hands-on exercise in Jupyter Labs to build a **fair and transparent and ethical** ML model for detecting fraudulent accounts * Discussion of the implications of module building for regulatory questions, reputation, and external costs * Final networking opportunity / happy hour |

**Module 4 Learning Objectives & Module Outline:**

*Course description*

* A module focused on the ethical questions surround the use of data to build ML models for cybersecurity. The module will introduce current conceptualizations of both fairness and privacy, discuss techniques for understanding and explaining models, and build an understanding for the many implications of using ML models for cybersecurity.

*Learning objectives*:

* Understand how and when ethical issues arise in using data to build ML models for cybersecurity
* Learn about key fairness considerations and debates over defining fairness
* Understand the implications of legal, philosophical, and technical conceptualizations of privacy for data collection, data retention, and model building
* Gain an appreciation for the personhood of data subjects and what this implies for data lifecycles
* Have hands-on experience trying to build an ML model for detecting fraudulent accounts where the straightforward solutions lead to profound ethical issues

*Asynchronous video outline*:

* Asynchronous video 1: Data cleaning and handling missing data (~5 minutes)
* Asynchronous video 2: Fairness in cybersecurity ML models (~10 minutes)
* Asynchronous video 3: Transparent and explainable ML models for cybersecurity (~5 minutes)
* Asynchronous video 4: Defining and actualizing privacy (~10 minutes)
* Asynchronous video 5: Collecting data from human subjects (~5 minutes)
* Asynchronous video 6: The pitfalls of anonymization (~5 minutes)
* Asynchronous video 7: Responsible data lifecycles (~5 minutes)

**Module 4: Synchronous Schedule (via Zoom):**

|  |  |  |
| --- | --- | --- |
| **Topic** | **Time** | **Notes:** |
| Introduction/  Core Concepts Recap | 15 min | * Introduction * Recap of basic concepts in videos * Questions about videos |
| Discussion of privacy in context | 15 min | * Focus on how philosophical definitions of privacy apply to cybersecurity challenges * Discuss whether differential privacy is a panacea (…it’s not) |
| Introduction to case study/group work | 5 min |  |
| Simulation/  Case study/Group Work | 30 min | * Hands-on lab about building a model to detect fraudulent accounts leveraging virtual case study (Jupyter Notebooks) * Divide students into Zoom breakout rooms * Pair students based on skill levels |
| Break | 10 min |  |
| Discussion | 15 min | * Discuss the externalities and implications of errors in ML models for cybersecurity from regulatory actions to bad press to increased calls to the help desk to issues with customer retention |
| Course Wrap up | 15 min |  |
| Networking Opportunity | 15 min | * TBD Virtual Happy Hour /speed meet-a-thon * Potentially led by class facilitator? |