Key Concepts for Future AISec Learners

Brent Lagesse and Colleen Lewis

Background and Overview

In 2021, a team from the University of Washington Bothell (UWB) and University of Illinois Urbana Champaign (UIUC) were awarded a grant entitled "Collaborative Research: EAGER: SaTC-EDU: Artificial Intelligence-Enhanced Cybersecurity: Workforce Needs and Barriers to Learning" to study workforce development for future jobs at the intersection of AI and Security. As part of this grant, the team interviewed experts working at the intersection of AI and Security to learn from them what skills and experiences are most important for new team members working in the field. The remainder of this document details the key concepts learned from industry. This information was also synthesized into a course at UWB that was taught in Winter 2023.

The intersection of AI and Security is an emerging area of both research and practice. Work in this area often focuses on two main approaches, AI for Security and Security for AI. Rather than separating these two concepts, this document treats the approaches as part of a broader approach. The development of an AI-based software product leads to the AI components of the software as an attack surface, so naturally any course that addresses AI for security needs to address Security for AI.

Key Concepts

Application Lifecycle is a multi-phase process that continues throughout the entire support-period of the application. This extends to both the AI and security aspects of the application.
End-to-End Development covers every stage required to develop the system.
Defense-in-Depth describes how systems are secured through multiple layers of defense in case a defense mechanism fails. This includes defense against Adversarial Machine Learning, which is a relatively new attack surface that explicitly targets AI-based systems.
Explainable AI is an AI system where the decision making process used by the AI can be understood by humans.
Usable Security is the idea that security mechanisms should be designed so that users understand what choices are in their best interest and do not disable security mechanisms.

Key Concepts with Supporting Fundamentals

Application Lifecycle is a multi-phase process that continues throughout the entire support-period of the application. This extends to both the AI and security aspects of the application.

- AI needs to be incorporated into the traditional software development lifecycle (SDLC) and Secure Development Lifecycle (SDL).
- AI as an attack surface needs to be considered as part of phases such as risk assessment, threat modeling, security testing, and security assessment.
- AI performance, not just in terms of predictive power, but also including computational performance, needs to be addressed during the SDLC. This should explicitly consider a plan for mitigating bias and continually testing for it.

End-to-End Development covers every stage required to develop the system.

- Data Collection must be done strategically to efficiently identify the correct information to collect, reduce data bias, and to collect data ethically.
- Preprocessing and feature extraction must incorporate domain knowledge from the problem domain in order to make best use of the data that is collected and avoid spending time and effort learning patterns that experts in the domain are already aware of.
- Model training and testing must account for security, bias, and explainability in addition to optimizing for general performance.
- Application development and deployment must be tested and analyzed for usability, security, and performance so that future development cycles can be improved and flaws can be corrected.

Defense-in-Depth describes how systems are secured through multiple layers of defense in case a defense mechanism fails. This includes defense against Adversarial Machine Learning which is a relatively new attack surface that explicitly targets AI-based systems.

- Best practices as part of a secure development lifecycle should be adopted for the development of the system.
- Best practices for deploying secure systems should be adopted and the system should be regularly assessed for security vulnerabilities and updated accordingly.
- Defenses should consider poisoning, evasion, inference, and extraction attacks.
- There are opportunities to deploy defenses against these attacks during data collection, preprocessing, model training, and system deployment.

Explainable AI is an AI system where the decision making process used by the AI can be understood by humans.

- Positive results on test data are insufficient for the deployment of a system. The system should be analyzed to understand why the results are positive.
- Explainable AI tools and techniques can be used to identify models that have learned from spurious correlations and bias in the training data.

Usable Security is the idea that security mechanisms should be designed so that users understand what choices are in their best interest and do not disable security mechanisms.

- Humans are involved at every stage. Users are diverse in their use cases, expertise, willingness to accept new tools, and many other ways. Their needs and preferences must be understood to improve adoptions of AI-enhanced security applications.
- Software should be designed using best practices usable security including principles such as Path of Least Resistance, Appropriate Boundaries, Explicit Authority, Visibility, Revocability, Expected Ability, Trusted Path, Identifiability, Expressiveness, and Clarity.
- User studies should be conducted to ensure that users operate the software in the intended manner.
- The impacts of False Positives, False Negatives, and Explainability on usability and user adoption of the software should be considered.