



Machine Learning Security, Privacy and Fairness Workshop - Feb 28th 08:30 to 1:00 pm

Facilitator: Dr. Mikel Rodriguez

Artificial intelligence and machine learning are being deployed across a growing number of real-world applications, including defense, transportation, and medicine. When it comes to AI systems in safety-critical contexts, significant challenges remain. In particular, all known machine learning algorithms are vulnerable to a growing number of privacy and security vulnerabilities. This workshop will provide an introduction to deep learning methods with an emphasis on understanding and improving their security, privacy, and fairness properties. The workshop will cover the basics of machine learning and introduce popular deep learning methods. It will focus on recently discovered vulnerabilities in state-of-the-art machine learning algorithms, their susceptibility to adversarial manipulation, and techniques for making machine learning robust to adversarial manipulation. It will cover state-of-the-art methods for explaining black-box deep learning models to enhance their transparency. It will also examine methods for machine learning that are designed to respect individual privacy and fairness.

Dr. Mikel Rodriguez manages the independent research and development efforts in the area of Machine Learning at the MITRE Corporation. For the past fifteen years, his main research has been on the exploration of how artificial intelligence (AI) and in particular Computer Vision can be used to help us solve problems for a safer world. He obtained his Ph.D. at UCF's Center for Research in Computer Vision. He was a visiting researcher at the Robotics Institute at Carnegie Mellon and a post-doctoral fellow at INRIA at the Département d'Informatique of Ecole Normale Supérieure in Paris, France. Dr. Rodriguez was the chair of the ODNI Video Analytics Research Working Group and is the senior technical adviser for the Pentagon's AI Team. He serves in the program committee for IEEE Computer Vision and Pattern Recognition, IEEE International Conference on Computer Vision, is an area chair of IEEE Transactions on Pattern Analysis and Machine Intelligence. His research is in the area of machine learning security and privacy.