Adversarial Artificial Intelligence – Dr. Chen Hajaj
Data Science and Artificial Intelligence Reseach Center, Ariel University

In recent years artificial intelligence (AI) research has had an increasing role in models and algorithms for security problems. Optimization methods and game-theoretic models of security (the Stackelberg security games in particular) have received special attention, in part because these models and associated tools have seen actual deployment in homeland security and sustainability applications.
In this talk I will focus on adversarial AI: while traditional AI ignores an entire class of important phenomena where relationships are adversarial, my talk will focus on a clash between system designer and a malicious adversary. This talk will touch two different problems: 1) a collection of potential assets that require protection, where a defender first allocates protection resources, and the attacker then responds with an optimal attack. 2) how robust are machine learning feature-based classifiers when facing realizable attacks.

Specifically, this talk will be based on the following papers:

- Chen Hajaj and Yevgeniy Vorobeychik. Adversarial task assignment. International Joint Conference on Artificial Intelligence and European Conference on Artificial Intelligence (IJCAI-ECAI 2018).
- Liang Tong, Bo Li, Chen Hajaj, Chaowei Xiao, and Yevgeniy Vorobeychik. Improving Robustness of ML Classifiers against Realizable, Evasion Attacks Using Conserved Feature. The 28th USENIX Security Symposium.