Special Session:

Trustworthiness of Machine Learning in Adversarial Environments

Yi Wang

Manhattan College, Riverdale, NY 10471

yi.wang@manhattan.edu

Aim and Scope

The Covid-19 pandemic have accelerated a transition towards an era of relying on cyberspace to intimately connect to the modern world with applications including smart transportation, smart manufacturing, smart healthcare, business, smart cities, modern power systems, social media, etc. However, these applications are vulnerable to cybersecurity attacks that adversely affect human's daily life. To defend against cybersecurity attacks, advanced artificial intelligence (AI) techniques have been leveraged.

In recent years, machine learning and deep learning algorithms have been the frontier of AI that reshape the current landscape of computing. However, AI system that are implemented by machine learning models suffer from adversarial attack vulnerability. Adversarial attacks aim at deceiving the AI system by inserting adversarial examples into the machine learning models to make false and/or inaccurate predictions. For example, with adversarial attacks, a panda can be falsely recognized as gibbon by adding adversarial perturbations in the machine learning model. In addition, a neural network with adversarial inputs can misclassify stop signs as speed limit 45 signs, and right turn signs as stop signs.

Adversarial attacks have been proven effective in adversarial environments in different applications. Rathore et al presented experiment show that adversarial attacks can achieve an average fooling rate of 98.68% against popular malware detection models. Additionally, adversarial attacks can target against face recognition and natural language processing in social media analysis. In addition, critical systems, such as power systems can be injected with adversarial examples to generate attacks that can be stealthy to bad data detector.

The aim of this special session is to establish a venue for scientists and engineers from academia, government, and industry to present and discuss latest advances and technologies on machine learning and deep learning theories and applications in adversarial environment settings.

The scope of this proposed special session is study and address the trustworthiness of machine learning and deep learning techniques applied in various applications under adversarial attacks, as well as a wide range of related issues of adversarial attacks and techniques to make machine learning, deep learning, and AI trustworthy. Thus, the scope of this special session is well aligned with the scope of the CyberSciTech 2022. The special session will seek original contributions, which address the key challenges and problems.

Topics

The topics of the special session include (but are not limited to):

- Adversarial Machine Learning and Reinforcement Learning
- Adversarial attacks and defenses in Internet of Things/Cyber-physical systems
- Adversarial attacks and defenses in software systems
- Adversarial attacks and defenses in malware detection and intrusion environments
- Data poisoning and evasion attacks
- Advanced techniques for generating adversarial examples
- Advanced defense mechanisms for adversarial attacks
- Vulnerability and security of machine learning/deep learning models
- AI assurance and security
- Secure machine learning systems in context of software security development
- New benchmark datasets for adversarial machine learning
- Industrial practices on adversarial machine learning and cybersecurity

Review Process

All the papers will be peer-reviewed by at least two technical program committee (TPC) members for this special session. Reviewers must provide at least 120 words comments in terms of synopsis, strength, weakness, and suggestions. TPC members will be recruited as soon as the proposal is approved.

Brief plans for dissemination

The organizers will disseminate one-page call for paper (CFP) brochure via social media, such as Linkedin and organizers' personal websites. TPC members will be asked to disseminate the CFP to their contacts as well. A few authors of recent published papers in adversarial learning and cybersecurity will be contacted via email to encourage them to submit a paper for this special session.

Submission Guidelines

This special session will be held in the 7th IEEE CyberSciTech 2022, September 12-15, Farlena, Italy. All papers should be prepared according to the CyberSciTech 2022 submission policy and should be submitted using the conference website (http://cyber-science.org/2022/cyberscitech/papersubmissions/). Paper submitted to this session will be 4-6 pages using the IEEE conference proceeding format:

http://www.ieee.org/conferences_events/conferences/publishing/templates.html

All papers accepted in this special session will be included in the CyberSciTech 2022 conference proceedings published by IEEE.

Important Dates

Submission Due: June 1, 2022

Author Notification Due: July 1, 2022

Camera-ready Submission: July 15, 2022

Organizer

Prof. Yi Wang,

Department of Electrical and Computer Engineering, Manhattan College, USA

yi.wang@manhattan.edu

Prof. Miaomiao Zhang,

Department of Computer Science, Manhattan College, USA

miaomiao.zhang@manhattan.edu

Prof. Bingyang Wei,

Department of Computer Science, Texas Christian University, USA

b.wei@tcu.edu

Reference

[1]. Chen, B., Ren, Z., Yu, C., Hussain, I., & Liu, J. (2019). Adversarial examples for cnn-based malware detectors. IEEE Access, 7, 54360-54371.

[2]. Rathore, H., Samavedhi, A., Sahay, S. K., & Sewak, M. (2021). Robust Malware Detection Models: Learning from Adversarial Attacks and Defenses. Forensic Science International: Digital Investigation, 37, 301183.

[3]. Vakhshiteh, F., Ramachandra, R., & Nickabadi, A. (2020). Threat of adversarial attacks on face recognition: A comprehensive survey. arXiv preprint arXiv:2007.11709.

[4]. Zhang, W. E., Sheng, Q. Z., Alhazmi, A., & Li, C. (2020). Adversarial attacks on deep-learning models in natural language processing: A survey. ACM Transactions on Intelligent Systems and Technology (TIST), 11(3), 1-41.

[5]. Tian, J., Wang, B., Wang, Z., Cao, K., Li, J., & Ozay, M. (2021). Joint Adversarial Example and False Data Injection Attacks for State Estimation in Power Systems. IEEE Transactions on Cybernetics.

Call for Papers

Special Session on Trustworthiness of Machine Learning in Adversarial Environments In conjunction with the 7th IEEE Cyber Science and Technology Congress (CyberSciTech 2022)

In recent years, machine learning and deep learning algorithms have been the frontier of Artificial Intelligence (AI) that reshape the current landscape of computing, which have achieved huge success in various domains, including smart transportation, smart manufacturing, smart healthcare, business, smart cities, modern power systems, social media, etc. However, AI system that are implemented by machine learning models suffer from adversarial attack vulnerability. Adversarial attacks aim at deceiving the AI system by inserting adversarial examples into the machine learning models to make false and/or inaccurate predictions. The aim of this special session is to establish a venue for scientists and engineers from academia, government, and industry to present and discuss latest advances and technologies on adversarial machine learning theories and applications, and related cyber security issues. The scope of this proposed special session is study and address adversarial machine learning techniques used in dealing with cybersecurity issues in various applications, as well as a wide range of related issues from machine learning, deep learning, AI, and cybersecurity in the following list of topics:

- Adversarial Machine Learning and Reinforcement Learning
- Adversarial attacks and defenses in Internet of Things/Cyber-physical systems
- Adversarial attacks and defenses in software systems
- Adversarial attacks and defenses in malware detection and intrusion environments
- Data poisoning and evasion attacks
- Advanced techniques for generating adversarial examples
- Advanced defense mechanisms for adversarial attacks
- Vulnerability and security of machine learning/deep learning models
- AI assurance and security
- Secure machine learning systems in context of software security development
- New benchmark datasets for adversarial machine learning
- Industrial practices on adversarial machine learning and cybersecurity

You are invited to submit a 4-6 pages original paper according to the CyberSciTech 2022 submission policy using the conference website. This special session will be held in the 7th IEEE CyberSciTech 2022, September 12-15, Farlena, Italy. All papers accepted in this special session will be included in the CyberSciTech 2022 conference proceedings published by IEEE. See details of submission policy via http://cyber-science.org/2022/cyberscitech/papersubmissions/.

Important Dates

Paper Submission:	**June 1, 2022**
Author Notification:	**July 1, 2022**
Camera- ready Submission:	**July 15, 2022**

Program Chair

Prof. Yi Wang	Manhattan College, USA	yi.wang@manhattan.edu

Program Cochairs:

Prof. Miaomiao Zhang	Manhattan College, USA	miaomiao.zhang@manhattan.edu
Prof. Bingyang Wei	Texas Christian University, USA	b.wei@tcu.edu