Enhancing AI Robustness in Cybersecurity: Challenges and Strategies

About this Research Topic

Submission closed

Background

Artificial intelligence systems are becoming increasingly pivotal in cybersecurity, yet their adoption is fraught with significant challenges. Current deployments often suffer from inadequate robustness evaluations, largely due to the limited availability and quality of datasets. This lack of thorough validation introduces various risks, including ethical dilemmas, legal issues, and violations of digital rights. Furthermore, as AI technologies orchestrate and manage network operations, they open up new avenues for attackers, exacerbating vulnerabilities inherent in AI and

Published in

Frontiers in Computer Science
Computer Security

2.7 impact factor
5.3 citescore

encounter within the cybersecurity domain. Specific emphasis will be placed on understanding adversarial AI attacks, such as injection tactics, and the defense mechanisms like adversarial training that can fortify systems against such exploits. The goal is also to delve into deception techniques, including the use of honeypots, digital twins, and virtual personas, and the role of explainable AI in enhancing transparency and trust in AI functionalities.

To adequately encapsulate the complexities at the intersection of AI and cybersecurity, this Research Topic will focus on both "AI for Cybersecurity" and "Cybersecurity for AI" aspects. Within this framework, the scope is clearly delineated as follows:

We primarily focus on the intersection of emerging AI technologies with the dynamic realm of cybersecurity threats.
We welcome contributions on a variety of pertinent themes:

- Economic implications of adversarial AI
- Ethical considerations in adversarial AI
- AI and ML techniques in cyber threat intelligence
- Machine learning in automated software testing
- Human factors affecting adversarial AI
- Defensive strategies against adversarial ML attacks
- Machine learning applications in analyzing cryptographic protocols
- Privacy-enhancing techniques in machine learning

This Research Topic is supported by the following EU-funded projects:

- AI-ASsisted cybersecurity platform empowering SMEs to defend against adversarial AI attacks (AIAS)
- Cloud-based Platform-agnostic Adversarial AI Defence framework (CPAID)
- AI Attack and Defense for the Smart Healthcare (ANTIDOTE)
- Revolutionised Enhanced Supply Chain Automation with Limited Threats Exposure (RESCALE)

Keywords: Adversarial AI attacks, Adversarial training, Cyber attacks detection and mitigation, Deception mechanism, AI in cybersecurity, Defensive strategies, Explainable AI, Robustness evaluation

Important note: All contributions to this Research Topic must be within the scope of the section and journal to which they are submitted, as defined in their mission statements. Frontiers reserves the right to guide an out-of-scope manuscript to a more suitable section or journal at any stage of peer review.

Topic editors

Apostolis Zarras
Foundation for...
Heraklion, Greece

Christos Xenakis
University of Piraeus
Piraeus, Greece

Christoforos Ntantogian
Department of Informati...
Corfu, Greece

Articles

See all articles (2) >

semantic layer of large language models

Yi Zhang · Jantan Aman

doi 10.3389/fcomp.2025.1683495

robustness against white-box adversarial examples

Ruben Stenhuis · Dazhuang Liu · Yanqi Qiao · Mauro Conti · Manos Panaousis · Kaitai Liang

doi 10.3389/fcomp.2025.1631561