**1. Adversarial AI Attacks on Next-Generation IDS/IPS in Critical Infrastructure**

**Background**
Modern critical infrastructure systems increasingly rely on AI-powered intrusion detection and prevention systems (IDS/IPS) to protect against cyber threats. However, these systems are vulnerable to adversarial attacks that can manipulate AI decision-making processes. With the growing sophistication of threat actors targeting power grids, water facilities, and transportation networks, understanding and mitigating adversarial vulnerabilities becomes crucial for national security and public safety.

Recent studies show that traditional signature-based detection can be easily bypassed by adversarial techniques. This project aims to investigate how Generative Adversarial Networks (GANs) and other AI-based methods can be used to create sophisticated attacks against next-generation IDS/IPS systems, while simultaneously developing robust defense mechanisms.

**Data**

* NSL-KDD and CICIDS2017 datasets for network intrusion detection
* SCADA network traffic datasets from critical infrastructure simulations (e.g. SWaT (Secure Water Treatment) Dataset, HAI 1.0: HIL-based Augmented ICS Security Dataset)
* Industrial control system (ICS) attack datasets including Stuxnet variants

**Research questions**

* How effective are GAN-generated adversarial samples in bypassing state-of-the-art AI-based IDS/IPS systems?
* What defense mechanisms (adversarial training, input sanitization, ensemble methods) provide the most robust protection against adversarial attacks?
* How do adversarial attacks perform against different types of critical infrastructure networks (power grid SCADA vs. water treatment systems)?
* Can real-time adversarial detection be implemented without significantly impacting system performance?

**Prerequisites**

* Strong knowledge of machine learning, deep learning, and GANs
* Understanding of network security and intrusion detection systems
* Programming skills in Python, TensorFlow/PyTorch
* Basic knowledge of industrial control systems and SCADA protocols

**Contact and application**
Grzegorz.Kolaczek@pwr.edu.pl

1. Y. Li, "A Review of the Impact of Adversarial Attacks on Intrusion Detection Systems Based on Machine Learning and Deep Learning," Advances in engineering technology research, 2025. doi: 10.56028/aetr.14.1.1601.2025
2. S. Saini, A. Chennamaneni, and B. A. Sawyerr, "A Review of the Duality of Adversarial Learning in Network Intrusion: Attacks and Countermeasures," arXiv preprint, 2024. doi: 10.48550/arxiv.2412.13880
3. K. Usha Rani and S. Dhingra, "A Comprehensive Survey of Deep Learning-Based Adversarial Network Intrusion Detection Systems," Proceedings, 2025. doi: 10.1109/icirca65293.2025.11089572
4. Abisoye, Ajayi, et al. "Using AI and machine learning to predict and mitigate cybersecurity risks in critical infrastructure." International Journal of Engineering Research and Development 21.2 (2025): 205-224.
5. Bushigampala, Bharath Kumar, and Anil Chowdary Inaganti. "Threat Detection in Critical Infrastructure Using AI Models." *International Journal of Acta Informatica* 2.1 (2023): 196-208.
6. Goffer, Mohammad Abdul, et al. "AI-Enhanced Cyber Threat Detection and Response Advancing National Security in Critical Infrastructure." *Journal of Posthumanism* 5.3 (2025): 1667-1689.
7. Abisoye, Ajayi, et al. "Using AI and machine learning to predict and mitigate cybersecurity risks in critical infrastructure." *International Journal of Engineering Research and Development* 21.2 (2025): 205-224.
8. Sarker, Iqbal H. "Multi‐aspects AI‐based modeling and adversarial learning for cybersecurity intelligence and robustness: A comprehensive overview." *Security and Privacy* 6.5 (2023): e295.

**2. Privacy-Preserving Federated Learning for Cyber Threat Intelligence in Banking**

**Background**
Financial institutions face increasingly sophisticated cyber threats but are limited in their ability to share threat intelligence due to strict privacy regulations and competitive concerns. Traditional centralized approaches to fraud detection require data sharing that violates customer privacy and regulatory compliance requirements such as GDPR and PSD2.

Federated learning offers a promising solution by enabling collaborative model training without exposing raw customer data. However, standard federated learning approaches still present privacy risks through gradient leakage and inference attacks. This project addresses these challenges by implementing advanced privacy-preserving techniques including differential privacy and secure multi-party computation.

**Data**

* Synthetic financial transaction datasets (Privacy-preserving Bank Account Fraud dataset), e.g. Bank Account Fraud (BAF) Dataset Suite, PaySim Synthetic Dataset)
* Credit card fraud detection datasets (IEEE-CIS Fraud Detection)
* Simulated multi-institutional federated learning scenarios with non-IID data distributions (AMLSim Dataset)

**Research questions**

* How can differential privacy be optimally calibrated to maximize model utility while ensuring strong privacy guarantees in financial fraud detection?
* What techniques best address the challenge of non-IID data distribution across different financial institutions?
* How does the privacy-utility trade-off compare between different secure aggregation protocols (secure multi-party computation vs. homomorphic encryption)?
* Can federated learning models achieve comparable accuracy to centralized approaches while maintaining regulatory compliance?

**Prerequisites**

* Strong background in machine learning and federated learning frameworks
* Knowledge of cryptography and privacy-preserving techniques
* Understanding of financial systems and regulatory requirements
* Programming experience with PySyft, TensorFlow Federated, or similar frameworks

**Contact and application**
Grzegorz.Kolaczek@pwr.edu.pl

1. Saura, Pablo Fernández, et al. "Privacy-preserving cyber threat information sharing leveraging fl-based intrusion detection in the financial sector." *International Workshop on Digital Sovereignty in Cyber Security: New Challenges in Future Vision*. Cham: Springer Nature Switzerland, 2022.
2. Sleem, A., and I. Elhenawy. "Enhancing Cyber Threat Intelligence Sharing through a Privacy-Preserving Federated Learning Approach." *Journal of Cybersecurity and Information Management* 2 (2022): 51-1.
3. Sakhare, Nitin N., et al. "A Decentralized Approach to Threat Intelligence using Federated Learning in Privacy-Preserving Cyber Security." *Journal of Electrical Systems* 19.3 (2023).
4. Timofte, Edi Marian, et al. "Federated Learning for Cybersecurity: A Privacy-Preserving Approach." *Applied Sciences* 15.12 (2025): 6878.
5. Di Salvo, Rosa, and Antonino Galletta. "Security and privacy concerns in Federated Learning systems: a systematic review." (2025).
6. Yuxin, Mao, and Wang Honglin. "Federated learning based on data divergence and differential privacy in financial risk control research." *Computers, Materials and Continua* 75.1 (2023): 863-878.
7. Wang, Fayao, et al. "Privacy-preserving robust federated learning with distributed differential privacy." *2022 IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*. IEEE, 2022.
8. Boscher, Cédric, et al. "Personalized Privacy-Preserving Federated Learning." *Proceedings of the 25th International Middleware Conference*. 2024.

**3. Cyber Deception and Honeypot-Based Defense Against Advanced Persistent Threats**

**Background**
Advanced Persistent Threats (APTs) represent one of the most challenging cybersecurity problems, characterized by sophisticated, long-term campaigns that evade traditional security measures. These attacks often involve nation-state actors or well-funded criminal organizations that can adapt their tactics to bypass conventional defenses.

Cyber deception technologies, including honeypots, honeynets, and decoy systems, offer a proactive defense approach by creating false targets that attract and trap attackers. This project focuses on designing intelligent deception mechanisms that can detect APT activities, study attacker behavior, and slow down attack progression while gathering valuable threat intelligence.

**Data**

* APT attack datasets (APT1, Lazarus Group, Fancy Bear campaign data)
* Honeypot interaction logs from Cowrie, Dionaea, and custom honeypot deployments (Medium-interaction SSH Honeypot Data, LLM Honeypot Logs)
* MITRE ATT&CK framework mappings for APT tactics, techniques, and procedures

**Research questions**

* What honeypot configurations and deception strategies are most effective at attracting and retaining APT attackers?
* How can machine learning be used to automatically adapt deception environments based on observed attacker behavior?
* What is the optimal integration approach between honeypot systems and existing SIEM/SOAR platforms for automated threat response?
* How do different types of deception (network-level, application-level, data-level) impact APT detection rates and false positive rates?

**Prerequisites**

* Strong understanding of cybersecurity principles and APT attack methodologies
* Knowledge of network security, system administration, and virtualization technologies
* Experience with security information and event management (SIEM) systems
* Programming skills in Python, bash scripting, and security automation tools

**Contact and application**
Grzegorz.Kolaczek@pwr.edu.pl

1. Aggarwal, Palvi, et al. "Decoys in cybersecurity: an exploratory study to test the effectiveness of 2-sided deception." *arXiv preprint arXiv:2108.11037* (2021).
2. Arora, Anuj. "Detecting and Mitigating Advanced Persistent Threats in Cybersecurity Systems."
3. Ayinde, Abiodun. *Evaluating the Performance of Deep Learning Algorithms in Detecting Advance Persistent Threats*. Diss. National University, 2025.
4. Morić, Zlatan, Vedran Dakić, and Damir Regvart. "Advancing Cybersecurity with Honeypots and Deception Strategies." *Informatics*. Vol. 12. No. 1. MDPI AG, 2025.
5. Hernández-Rivas, Adrián, Víctor Morales-Rocha, and J. Patricia Sánchez-Solís. "Towards autonomous cybersecurity: A comparative analysis of agnostic and hybrid AI approaches for advanced persistent threat detection." *Innovative Applications of Artificial Neural Networks to Data Analytics and Signal Processing*. Cham: Springer Nature Switzerland, 2024. 181-219.
6. Yusof, Zainuddin Bin. "Exploration of advanced persistent threats: techniques, mitigation strategies, and impacts on critical infrastructure." *International Journal of Advanced Cybersecurity Systems, Technologies, and Applications* 8.12 (2024): 1-9.
7. Sharma, Amit, et al. "Advanced persistent threats (apt): evolution, anatomy, attribution and countermeasures." *Journal of Ambient Intelligence and Humanized Computing* 14.7 (2023): 9355-9381.
8. Khalid, Mohd Nor Akmal, Amjed Ahmed Al-Kadhimi, and Manmeet Mahinderjit Singh. "Recent developments in game-theory approaches for the detection and defense against advanced persistent threats (APTs): a systematic review." *Mathematics* 11.6 (2023): 1353.

**4. Explainable AI and Digital Twins for Real-Time Anomaly Detection in Industrial Systems**

**Background**
Industrial control systems are increasingly targeted by cyber-physical attacks that can cause physical damage and safety hazards. Traditional anomaly detection methods often lack the context of physical processes and provide limited explanability, making it difficult for operators to understand and respond to threats effectively.

Digital twin technology combined with explainable AI offers a promising approach to bridge the gap between cybersecurity and operational technology. By creating virtual replicas of industrial systems and using physics-informed neural networks, this project aims to develop anomaly detection systems that not only identify threats but also provide clear explanations of their decisions.

**Data**

* SWaT (Secure Water Treatment) testbed dataset
* WADI (Water Distribution) system attack dataset
* HAI (Hardware-in-the-loop Augmented Industrial) dataset
* SCADA system simulation data from industrial process simulators

**Research questions**

* How can physics-informed neural networks be effectively integrated with digital twin models for anomaly detection in industrial systems?
* What explainability techniques (SHAP, LIME, attention mechanisms) provide the most actionable insights for industrial operators?
* How does the performance of digital twin-based detection compare to traditional statistical and machine learning approaches?
* Can real-time explainable anomaly detection be achieved within the latency constraints of industrial control systems?

**Prerequisites**

* Strong background in machine learning and explainable AI techniques
* Knowledge of industrial control systems, SCADA protocols, and cyber-physical systems
* Understanding of physics-informed neural networks and digital twin concepts
* Programming skills in Python, MATLAB/Simulink, or industrial simulation tools

**Contact and application**
Grzegorz Kolaczek@pwr.edu.pl

1. Huang, Huiyue, et al. "Digital twin-driven online anomaly detection for an automation system based on edge intelligence." *Journal of Manufacturing Systems* 59 (2021): 138-150.
2. Moosavi, Sajad, et al. "Explainable AI in manufacturing and industrial cyber–physical systems: A survey." *Electronics* 13.17 (2024): 3497.
3. Bertalanič, Blaž, Vid Hanžel, and Carolina Fortuna. "Explainable semantic wireless anomaly characterization for digital twins." *Computer Networks* 251 (2024): 110660.
4. De Benedictis, Alessandra, et al. "Digital twins for anomaly detection in the industrial internet of things: Conceptual architecture and proof-of-concept." *IEEE Transactions on Industrial Informatics* 19.12 (2023): 11553-11563.
5. Oyedotun, Samuel Abiodun, Godfrey Perfectson Oise, and Chukwuma Emmanuel Ozobialu. "Towards Intelligent Cybersecurity in SCADA and DCS Environments: Anomaly Detection Using Multimodal Deep Learning and Explainable AI." *Journal of Science Research and Reviews* 2.3 (2025): 20-31.
6. Lalithadevi, B., and S. Krishnaveni. "A Comprehensive Survey on Enhancing Digital Twin Security Systems with Explainable AI Techniques." *2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*. IEEE, 2025.
7. Mikołajewska, Emilia, et al. "Generative AI in AI-based digital twins for fault diagnosis for predictive maintenance in Industry 4.0/5.0." *Applied Sciences* 15.6 (2025): 3166.
8. Bolat-Akça, Burcu, and Elif Bozkaya-Aras. "Digital twin-assisted intelligent anomaly detection system for Internet of Things." *Ad Hoc Networks* 158 (2024): 103484.