AI-ASsisted cybersecurity platform empowering SMEs to defend against adversarial AI attacks

AIAS NEWSLETTER

Issue 1 | August 2024

AI systems find applications in various technical fields. However, their adoption exposes early users to vulnerabilities, such as data corruption, model theft, and adversarial samples. The lack of tactical and strategic capabilities to defend, identify, and respond to attacks on these AI-based systems is a significant concern. Adversaries exploit this vulnerability, creating a new attack surface that specifically targets Machine Learning and Deep Learning systems, posing a substantial threat to critical sectors like finance and healthcare. Addressing these challenges, the MSCA-funded AIAS project aims to conduct research on adversarial AI and develop an innovative security platform for organisations. This platform will employ adversarial AI defence methods, deception mechanisms, and explainable AI solutions to empower security teams, fortifying AI systems against potential attacks.

PROJECT COORDINATION

Prof. Christos Xenakis
School of Information and Communication Technologies
Department of Digital Systems
University of Piraeus
Karaoli and Dimitriou 80,PC 18534, Piraeus, Greece
Tel: +30 210 4142776
email: xenakis@unipi.gr

PROJECT DETAILS

Project number: 101131292
Project Website: aias-project.eu
Project start: 1st January 20194
Duration: 48 Months
Total cost: EUR 1564000
EC Contribution: EUR 1564000

This project has received funding from the European Union under HORIZON-TMA-MSCA-SE, Topic HORIZON-MSCA-2022-SE-01-01, Grant Agreement No 101131292.

AIAS

AI-ASsisted cybersecurity platform empowering SMEs to defend against adversarial AI attacks

AIAS Motivation

- AI systems undiscovered vulnerabilities, such as data corruption, model theft and adversarial samples.
- Lack of tactical and strategic capabilities to defend, identify, and respond to attacks on their AI-based systems.
- Adversaries have created a new attack surface to exploit AI system vulnerabilities.
- Minimal to no prior robustness tests (i.e., against adversarial attacks) .
- Often bound to utilize datasets with limited quantity and quality.
- The scarce and often superficial vetting of AI technologies raises ethical, legal and digital rights concerns.

AIAS Approach

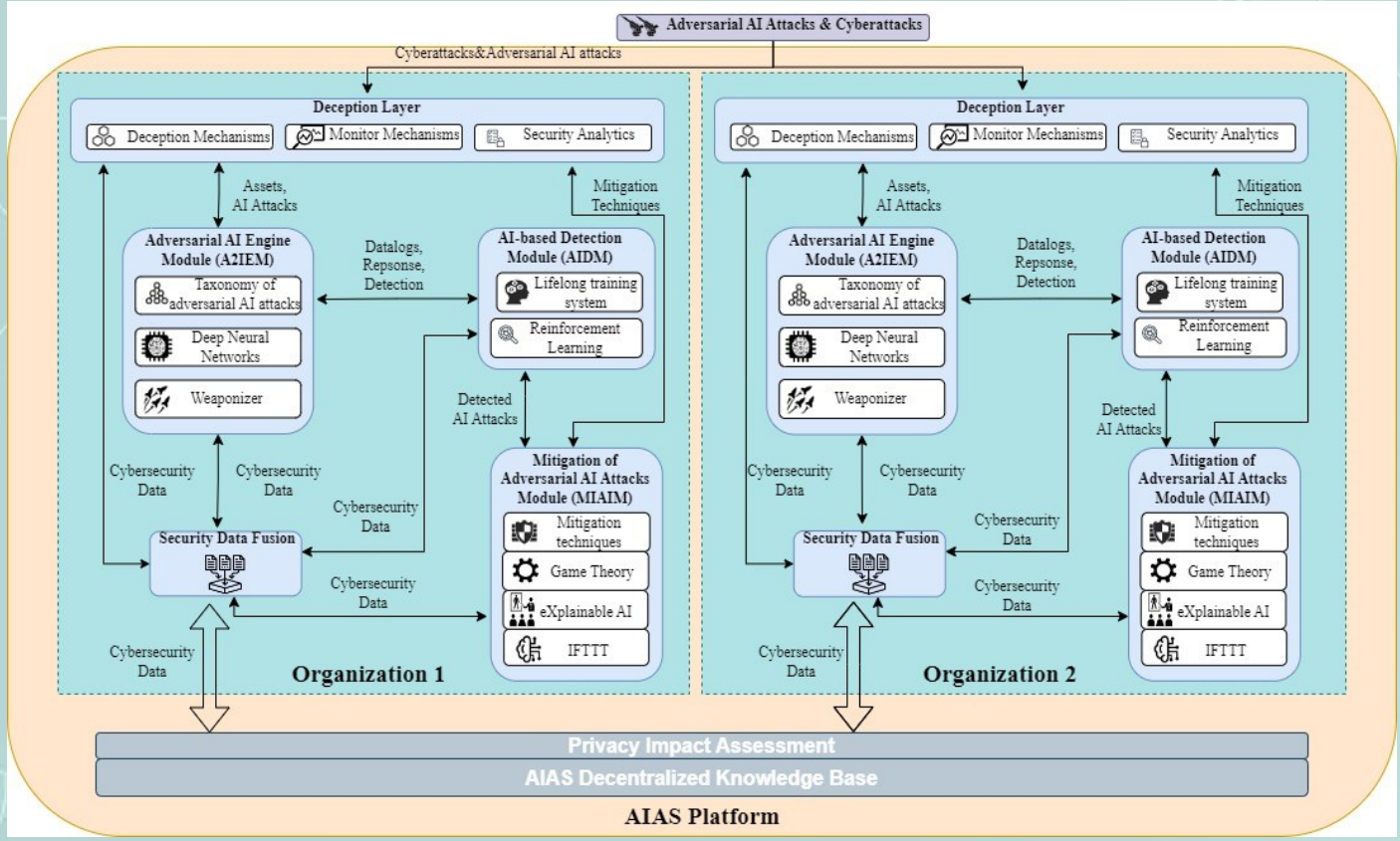
- AI systems undiscovered vulnerabilities, such as data corruption, model theft and adversarial samples.
- Lack of tactical and strategic capabilities to defend, identify, and respond to attacks on their AI-based systems.
- Adversaries have created a new attack surface to exploit AI system vulnerabilities.
- Minimal to no prior robustness tests (i.e., against adversarial attacks) .
- Often bound to utilize datasets with limited quantity and quality.
- The scarce and often superficial vetting of AI technologies raises ethical, legal and digital rights concerns.

This project has received funding from the European Union under HORIZON-TMA-MSCA-SE, Topic HORIZON-MSCA-2022-SE-01-01,

AIAS

AI-ASsisted cybersecurity platform empowering SMEs to defend against adversarial AI attacks

AIAS architecture overview



The AIAS platform consists of Six distinct modules.

- **Deception Layer** implements deception mechanisms based on high-interaction honeypots, digital twins and virtual personas to create a virtual imitation of the organization in order to deceive adversaries and lure potential attacks on the organization's AI models and AI-based systems.

This project has received funding from the European Union under HORIZON-TMA-MSCA-SE, Topic HORIZON-MSCA-2022-SE-01-01, Grant Agreement No 101131292.

AIAS

AI-ASsisted cybersecurity platform empowering SMEs to defend against adversarial AI attacks

AIAS architecture overview

- **Adversarial AI Engine Module (A2IEM)** generates adversarial AI attacks that will be deployed to test the robustness of the organizations AI systems.
- **AI-based Detection Module (AIDM)** will employ *Life-long Reinforcement Learning* to detect known and unknown adversarial AI attacks and cyberattacks. Life-long reinforcement learning's goal is to continuously and dynamically train the Reinforcement Learning model using the new information acquired from the ***Data Security Fusion base*** and the ***AIAS Decentralized Knowledge Base***.
- The **Mitigation of Adversarial AI Attacks Module (MIAIM)** will utilize the data from already detected attacks to recommend mitigation actions to human operators, based on comprehensible, transparent and explainable recommendations empowering security teams with a game theory-based and XAI recommendation engine for mitigation actions. A part of this module is the **Adversarial AI Mitigator (AIM)** that exploits game theory methods to collect and learn all possible mitigation techniques that can deal with each adversarial AI attack.
- **Security Data Fusion** base is a data pool that gathers all the security data regarding the attacks that have occurred in the organization and were captured by the deception mechanisms, the attacks that have been detected by the AI-based Detection Module, the generated adversarial AI attacks, as well as it preserves the deployed mitigation methods.

This project has received funding from the European Union under HORIZON-TMA-MSCA-SE, Topic HORIZON-MSCA-2022-SE-01-01, Grant Agreement No 101131292.

AIAS

AI-ASsisted cybersecurity platform empowering SMEs to defend against adversarial AI attacks

AIAS architecture overview

- **AIAS Decentralized Knowledge Base** *is* a Distributed Ledger Technology (DLT)-based InterPlanetary File System (IPFS) that collects all the security data from various instances of AIAS that have been installed in different organizations in a decentralized and distributed knowledge base.

News & Events

AIAS Project Kickoff meeting

AIAS secondment from FOGUS to CNIT

NEW AIAS Secondment from UPRC to BEIA

This project has received funding from the European Union under HORIZON-TMA-MSCA-SE, Topic HORIZON-MSCA-2022-SE-01-01, Grant Agreement No 101131292.

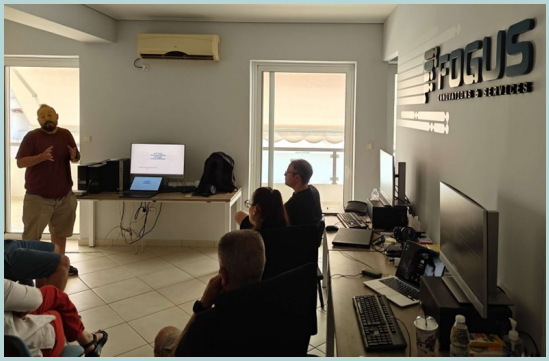
AIAS

AI-ASsisted cybersecurity platform empowering SMEs to defend against adversarial AI attacks

News & Events

[AIAS CO-organizes the 4th IWAPS 2024]

[Agenda for the 4th IWAPS]



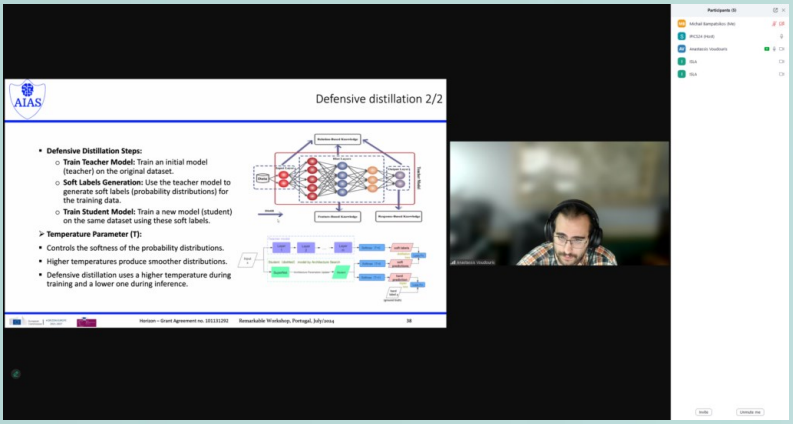
[Secondment from UPV to FOGUS]

[Thank you Nacho!]

[AIAS technical deep dive presentation in "AI Workshop 2024"]

[AIAS overall presentation in "AI Workshop 2024"]

[AIAS overall presentation in "AI Workshop 2024"]



This project has received funding from the European Union under HORIZON-TMA-MSCA-SE, Topic HORIZON-MSCA-2022-SE-01-01, Grant Agreement No 101131292.

AIAS Publication

- Petihakis, G., Farao, A., Bountakas, P., Sabazioti, A., Polley, J. and Xenakis, C., 2024, July. AIAS: AI-ASsisted cybersecurity platform to defend against adversarial AI attacks. In Proceedings of the 19th International Conference on Availability, Reliability and Security (pp. 1-7).

Upcoming Technical Deliverables

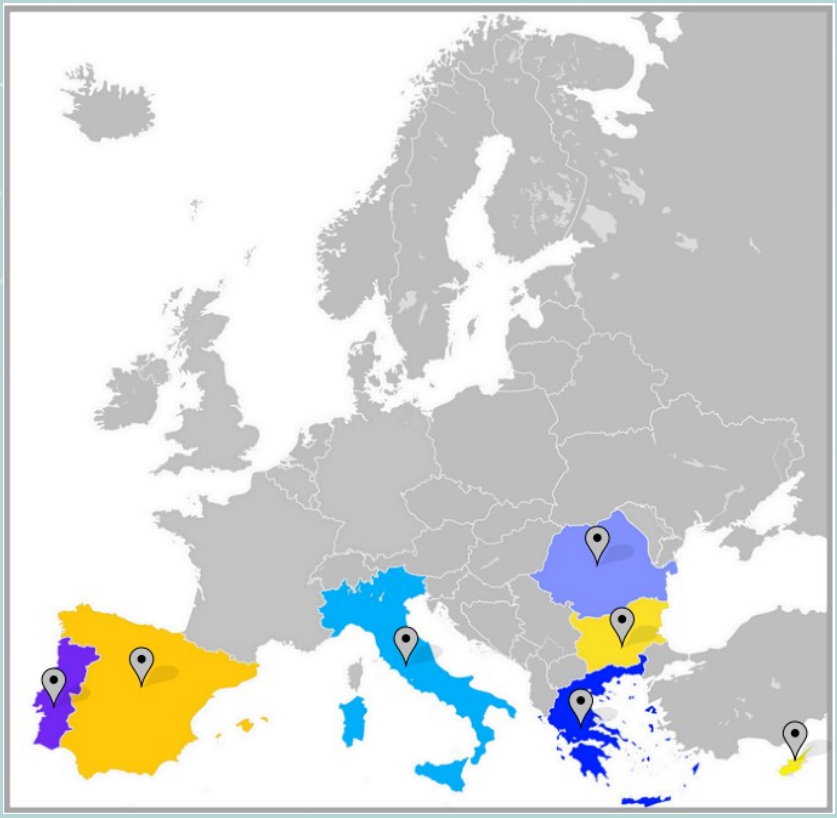
- D2.1-Requirements & Reference Architecture (December/2024)
- D2.2-Specification & Business cases (June/2025)
- D3.1-AIAS Deception Layer (August/2025)

This project has received funding from the European Union under HORIZON-TMA-MSCA-SE, Topic HORIZON-MSCA-2022-SE-01-01, Grant Agreement No 101131292.

AIAS

AI-ASsisted cybersecurity platform empowering SMEs to defend against adversarial AI attacks

Meet the Consortium



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS

cnit consorzio nazionale interuniversitario per le telecomunicazioni

UNIVERSIDAD DE MÁLAGA

K3Y R&D AND CYBER SECURITY

Beia CONSULT INTERNATIONAL

FOGUS INNOVATIONS & SERVICES

PDM

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Suite5 We Deliver Intelligence

Follow us for our latest news!

www.aias-project.eu

@AIAS.MSCA

@AIAS MSCA

@AIAS_MSCA

This project has received funding from the European Union under HORIZON-TMA-MSCA-SE, Topic HORIZON-MSCA-2022-SE-01-01, Grant Agreement No 101131292.