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

AIAS NEWSLETTER

Issue 2 | December 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 2024
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 Objectives

- **Holistic Protection:** Conceptualize and develop a service architecture integrating AI-empowered applications, deception mechanisms, and mitigation techniques towards the holistic protection of organizations against cyberattacks and adversarial AI.
- **Attack Scenarios:** Design and develop a novel adversarial AI engine for creating attack scenarios tailored to the characteristics of the targeted organisations' hardware and software infrastructure.
- **Novel Intelligent Deception Methods:** Design and implement novel intelligent deception methods based on high-interaction honeypots, digital twins, and virtual personas.
- **AI-based Methods for Protection:** Design, develop, and assess AI-based methods for the detection and mitigation of cyberattacks including adversarial AI attacks as well as conceptualize and implement data collection and fusion methods.
- **XAI-based Recommendation Engine:** Develop and verify explainable AI (XAI)-based recommendation engine empowering human-in-the-loop proactive decisions to thoroughly mitigate adversarial AI attacks.
- **Real-life Usage:** Assess the functionality, effectiveness and efficiency of AIAS in real-life scenarios.

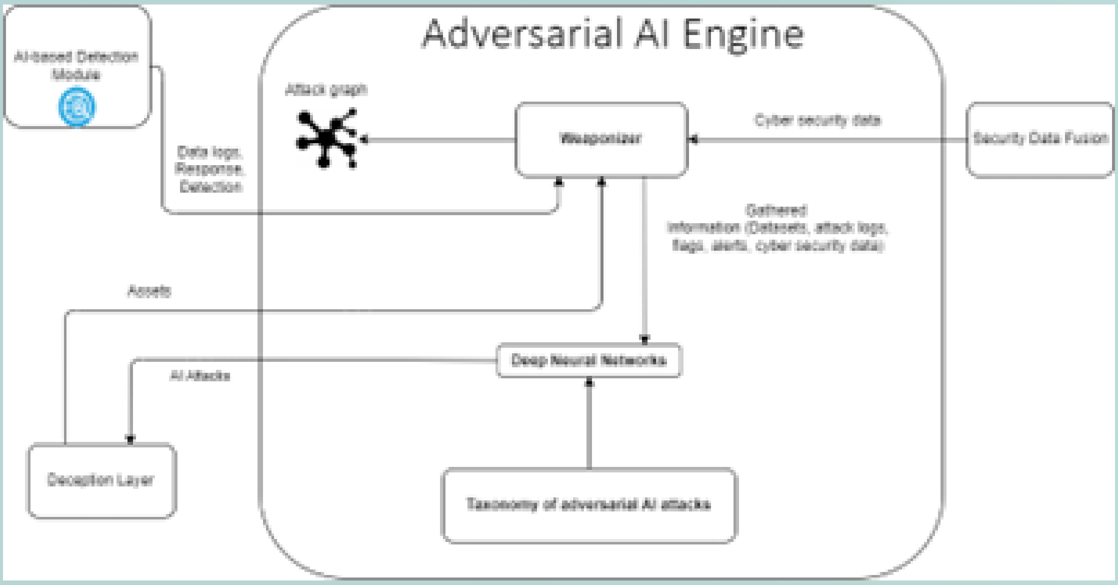
This project has received funding from the European Union under HORIZON-TMA-MSCA-SE, Topic HORIZON-MSCA-2022-SE-01-01, Grant Agreement No 101131292.

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

AIAS Architecture

- The AIAS architectural framework is constituted by an integrated set of components, each of which is designed to contribute to the formation of a unified cybersecurity defence system capable of safeguarding SMEs from sophisticated adversarial AI and cyber threats.
- The system comprises several key components, including the Adversarial AI Engine, the Deception Layer, the AI-based Detection Module, the XAI-based Mitigation Engine, and the Security Data Fusion and Decentralised Knowledge Base.

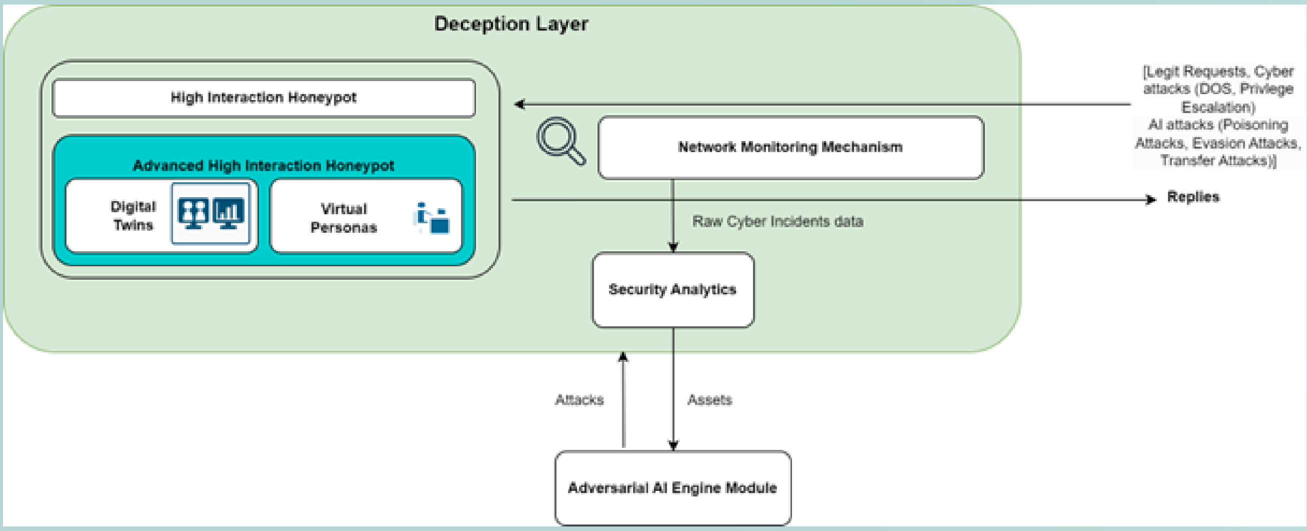
AIAS Adversarial AI Engine and Deception


Adversarial AI Engine
AI-based Detection Module
Attack graph
Weaponizer
Cyber security data
Security Data Fusion
Data logs, Response, Detection
Gathered information (Datasets, attack logs, flags, alerts, cyber security data)
Assets
Deep Neural Networks
AI Attacks
Deception Layer
Taxonomy of adversarial AI attacks

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.

The Adversarial AI Engine Module (AI2EM) is responsible for generating adversarial AI attacks, including Poisoning, Evasion, and Transfer attacks, which are strategically directed towards the AIAS deception layer and constructs Attack Graphs highlighting pathways through which various vulnerabilities might be exploited. AI2EM is composed of three core sub-modules: 1) The Weaponizer, 2) Deep Neural Networks (DNNs), and 3) A Taxonomy of Adversarial AI Attacks.

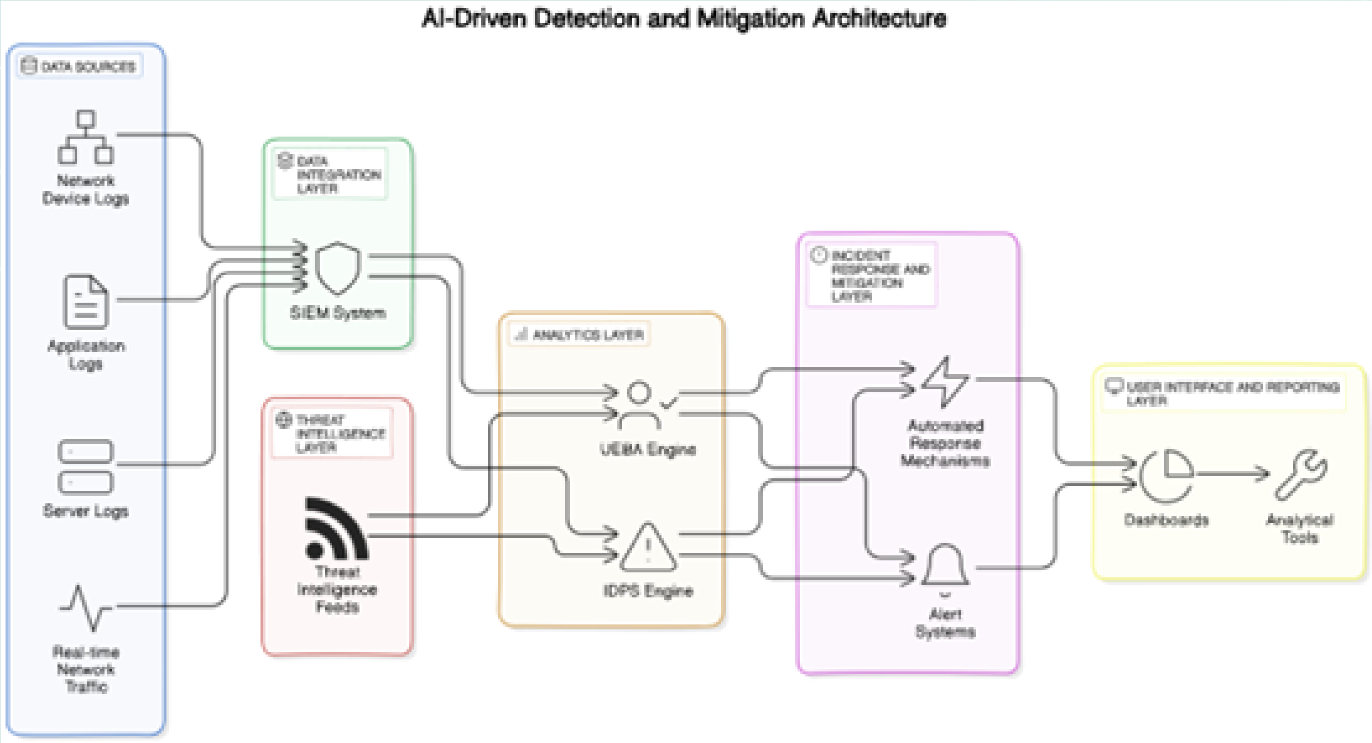
AIAS Deception Layer


Deception Layer
High Interaction Honeypot
Advanced High Interaction Honeypot
Digital Twins
Virtual Personas
Network Monitoring Mechanism
[Legit Requests, Cyber attacks (DOS, Privilege Escalation)
AI attacks (Poisoning Attacks, Evasion Attacks, Transfer Attacks)]
Replies
Raw Cyber Incidents data
Security Analytics
Attacks
Assets
Adversarial AI Engine Module

The Deception Layer constructs a virtual replica of the organization’s Information Communication Technology (ICT) infrastructure, designed to mislead adversaries and attract potential attacks on the organization’s AI models and AI-driven systems. The Deception Layer integrates a deception mechanism, utilizes security analytics to process data collected and deploys network monitoring tools.

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-driven Detection and Mitigation



AI-Driven Detection and Mitigation Architecture

DATA SOURCES
Network Device Logs
Application Logs
Server Logs
Real-time Network Traffic
DATA INTEGRATION LAYER
SIEM System
THREAT INTELLIGENCE LAYER
Threat Intelligence Feeds
ANALYTICS LAYER
UEBA Engine
IDPS Engine
INCIDENT RESPONSE AND MITIGATION LAYER
Automated Response Mechanisms
Alert Systems
USER INTERFACE AND REPORTING LAYER
Dashboards
Analytical Tools

The AI-driven detection and mitigation technologies include the User and Entity Behaviour Analytics (UEBA) and Intrusion Detection and Prevention Systems (IDPS). These technologies leverage artificial intelligence to detect anomalies and mitigate potential threats effectively. The architecture consists of the Data Sources, which include logs from applications, servers, network devices, and real-time network traffic.

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.

News & Events

AIAS CO-organizes 1st Plenary Meeting in Piraeus

Secondment from BEIA to UPRC

Thank you Maria!

Secondment from UPRC to BEIA

Thank you Athanasia!

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.

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

News & Events

AIAS participated in European Researchers’ Night 2024

AIAS Steals the Spotlight at Greek Researcher’s Night 2024

AIAS at Science Week 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

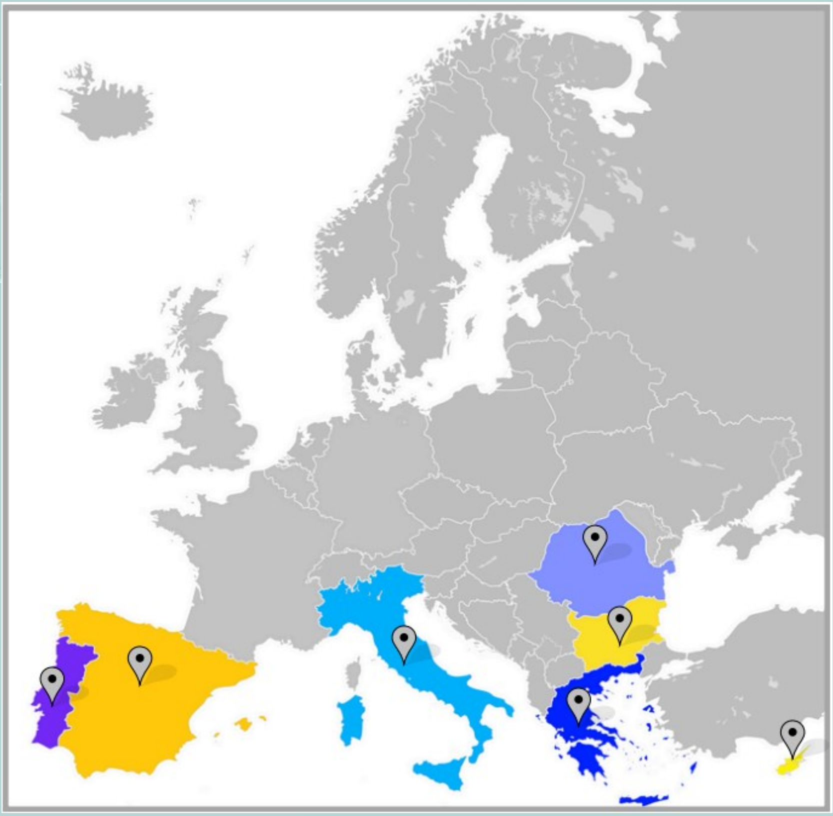
- D2.2-Specification & Business cases (June/2025)
- D3.1-AIAS Deception Layer (August/2025)

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.