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

AIAS NEWSLETTER

Issue 3 | April 2025

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.

AIAS

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.

Explainability and Human-Centric Decision Support

A fundamental tenet of the AIAS architectural framework is the provision of transparency and explainability in its recommendations. This is achieved through the deployment of an XAI-based Mitigation Engine. In light of the necessity for human involvement in cybersecurity, particularly in SMEs with constrained cybersecurity resources, the XAI component furnishes security operators with transparent and comprehensible justifications for recommended actions. The mitigation engine employs the use of SHAP and LIME techniques so as to elucidate the logic behind the mitigation suggestions that it makes.

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.

Security Data Fusion

Data gathering is an essential procedure for every AI-based task and several methods exist in the literature to collect data from a single or multiple sources. Web crawling and web scraping are two well-known methods that have been heavily deployed to create both small- and large-scale datasets.

AIAS will advance the state of the art by investigating and implementing a security data fusion approach that will combine data originating from different sources and data types. The security data fusion intents to combine IPFS with Hyperledger Fabric to create federated data storage. Moreover, AIAS envisages to design and develop an AI-based web crawler to automatically collect data from the web including also dark web regarding adversarial AI attacks and cyberattacks, vulnerabilities in AI systems as well as malevolent information about the organization.

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

New Publication Acknowledging AIAS: Trust Score Prediction in IoT Ecosystems

Secondment from FOGUS to CNIT

Thank you Georgios!

AIAS Leading the Way in AI Security at Cybersecurity Capacity-Building Workshop!

Secondment from UMA to FOGUS

Thank you Iman!

AIAS Supporting Cutting-Edge Research in IoT Trust Management!

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

Exciting Participation of AIAS at the 2nd International Exhibition Industry Tec 2025

AIAS at INDUSTRY.TEC Forum: Contributing to a Thought-Provoking Panel on Healthcare Digital Transformation

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 Supporting Cybersecurity Skills Development at the 12th Information Security Conference!]

[AIAS Co-Organizes the 5th IWAPS 2025 at ARES: Advancing Privacy-Preserving AI & Cybersecurity]

[AIAS at CodeWeek: Mentoring the Next Generation of Innovators]

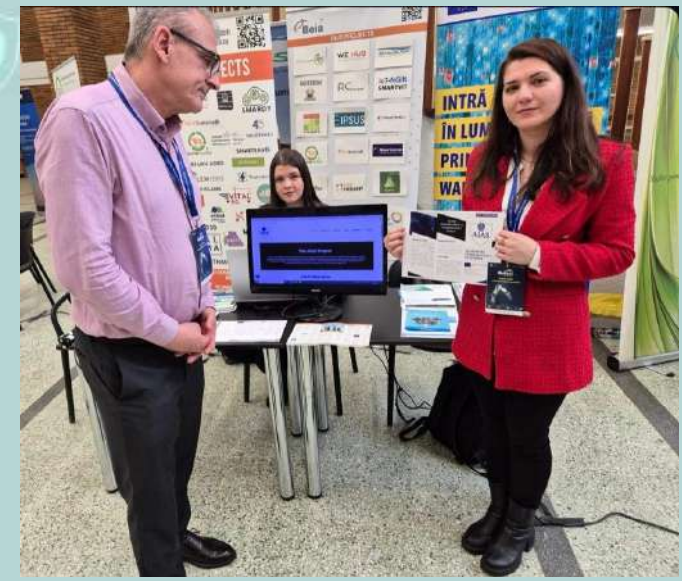
This project has received funding from the European Union under HORIZON-TMA-MSCA-SE, Topic HORIZON-MSCA-2022-SE-01-01, Grant Agreement No 101131292.

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

News & Events

Ignacio experience in Greece as secondment researcher. Thank you Ignacio!

AIAS-Sponsored Presentation on Zero-Knowledge Proofs



AIAS Showcased at POLI BIOFEST 2025 by BEIA Team

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 Presented at MEDFEST 2025]

[AIAS Engages Future Innovators at “Practica la ETTI”]

[AIAS Project Participates in Build with AI – Bucharest 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.

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

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).
- Bampatsikos M, Politis I, Ioannidis T, Xenakis C. Trust Score Prediction and Management in IoT Ecosystems Using Markov Chains and MADM Techniques. IEEE Transactions on Consumer Electronics. 2025 Jan 17.

Upcoming Technical Deliverables

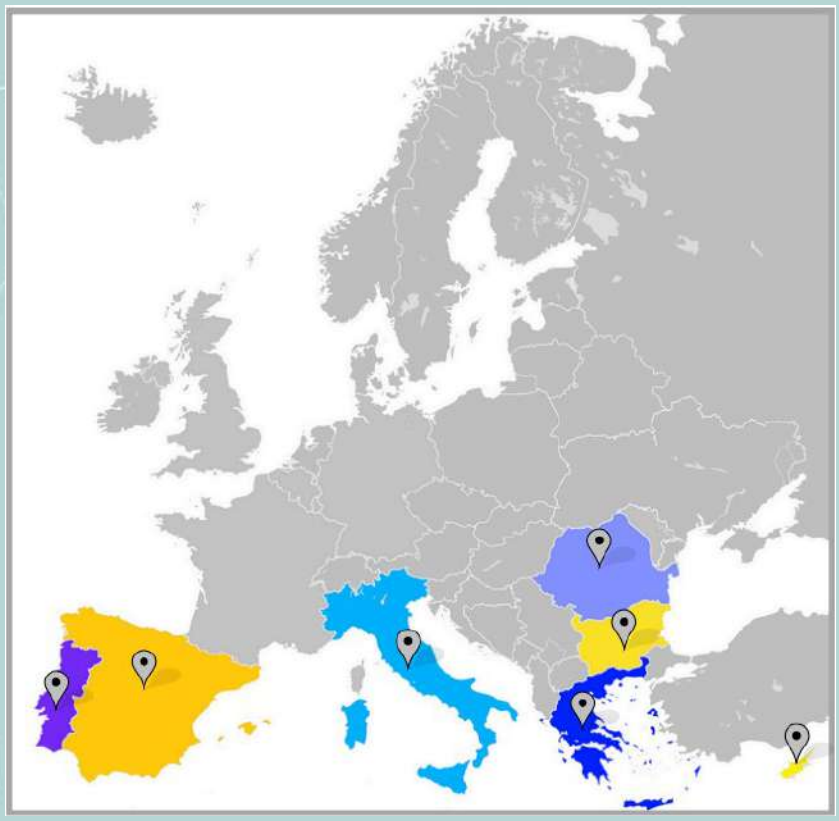
- D2.2-Specification & Business cases (June/2025)
- D3.1-AIAS Deception Layer (August/2025)

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.