"AI FOR CYBERSECURITY" & "CYBERSECURITY FOR AI"

MOTIVATION

The use of AI-based systems creates new attack surfaces for hosting frameworks and AI/ML algorithms. AI-based components are often black boxes, making them prime targets for threat actors who have developed techniques to impair the robustness of AI systems with adversarial AI attacks, violate the integrity of AI models, and bypass or disable the models by querying them with malicious input.

DESCRIPTION

AIAS aims to design and develop an innovative security platform to protect AI systems using adversarial AI defence, deception techniques, and explainable AI models, enhancing resilience against cyberattacks.

ACTIONS

AIAS project aims to perform in-depth research on adversarial AI to design and develop an innovative AI-based security platform for the protection of AI systems' technical robustness and AI-based operations of organizations, relying on Adversarial AI defence methods, deception mechanisms as well as on eXplainable AI solutions (XAI).

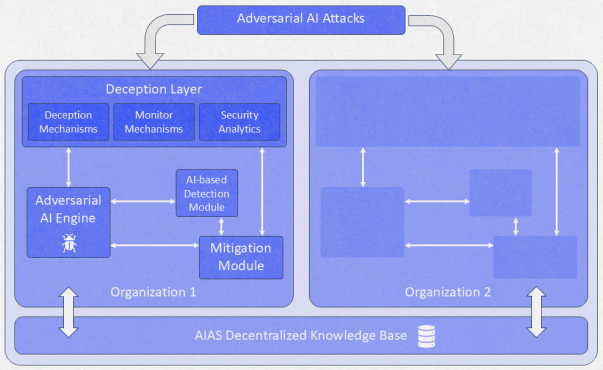
DECEMBER 2024

AIAS

AI-ASSISTED CYBERSECURITY PLATFORM

HORIZON-MSCA-2022-SE-01-01;
HORIZON.1.2 - Marie Skłodowska-Curie Actions (MSCA)

PROJECT WEBSITE: https://www.aias-project.eu/
PROJECT START: 1st January 2024
DURATION 48months
GRANT AGREEMENT: 101131292
EU CONTRIBUTION: EUR 1 564 000
COORDINATION: University of Piraeus Research Center (Greece)



IMPORTANCE OF AI IN AIAS

Adversarial AI Engine
Creation of attack scenarios adapted to the organization's characteristics.

Generative Adversarial Networks (GANs)
Detection of adversarial AI attacks.

XAI
Recommendation engine for further attack mitigation and comprehension of the decisions taken by the AI.

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 organizations' 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 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.

METHODOLOGY

Phase 1: System requirements and platform's main components identifications.

- Identify and define the security, privacy, functional and ethical requirements.
- Perform SOTA reviews in the key program areas, deception methods, AI-driven detection and mitigation methods.
- Specify the tools and applications that will be used for the implementation of each AIAS module.

Phase 2: Implementation and validation of the main platform components. Each module will be designed following well-known techniques, implementing related methods and guidelines provided by the EU, as well as using cutting-edge technologies.

Fase 3: Integration, proof-of-concept study and real-life assessment. The main objective in this Phase is to deliver the AIAS platform and the modules that comprise it shall be functional and seamlessly work. Once the platform integration is completed, all partners will assess the platform through carefully selected real-life pilot use cases. This task may include modifications of the platform based on the feedback acquired during the experiments.