“AI PENTRU SECURITATE CIBERNETICĂ” ȘI “CSECURITATE CIBERNETICĂ PENTRU AI”

MOTIVATIE

Utilizarea sistemelor bazate pe AI creează noi noi puncte de vulnerabilitate pentru infrastructurile de găzduire (hosting frameworks) și algoritmii AI/ML. Componentele bazate pe AI sunt adesea considerate “cutii negre” (black boxes), devenind ținte principale pentru actorii rău intenționați care au dezvoltat tehnici de compromitere a robusteții sistemelor AI prin atacuri adversariale, de încălcare a integrității modelelor de IA și de ocolire sau dezactivare a acestora prin interogări cu date malițioase.

DESCRIERE

AIAS își propune să proiecteze și să dezvolte o platformă inovatoare de securitate pentru protejarea sistemelor de inteligență artificială, folosind apărarea împotriva atacurilor adversariale (adversarial AI defence), tehnici de decepție (deception techniques) și modele de IA explicabile (explainable AI models), sporind reziliența împotriva atacurilor cibernetice.

ACȚIUNI

Proiectul AIAS își propune să efectueze cercetări aprofundate în domeniul inteligenței artificiale adversariale pentru a proiecta și dezvolta o platformă inovatoare de securitate bazată pe IA, destinată protejării robusteții tehnice a sistemelor de IA și a operațiunilor bazate pe IA ale organizațiilor. Platforma va utiliza metode de apărare împotriva atacurilor adversariale, mecanisme de decepție, precum și soluții explicabile de inteligență artificială (XAI).

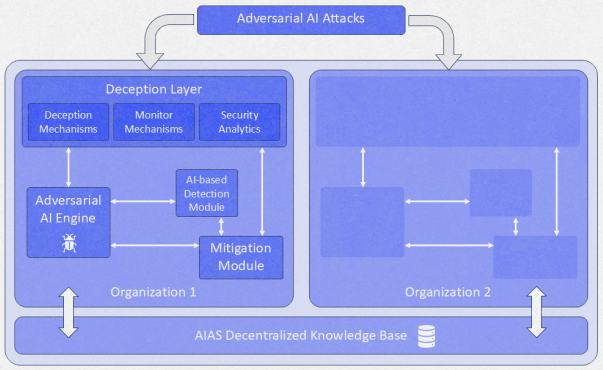
DECEMBRIE 2024

AIAS

PLATFORMA DE SECURITATE CIBERNETICĂ ASISTATĂ DE AI

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

WEBSITE PROJECT: https://www.aias-project.eu/
ÎNCEPTUL PROIECTULUI: 1 Ianuarie 2024
DURATA: 48 luni
GRANT AGREEMENT: 101131292
CONTRIBUTIE UE: EUR 1 564 000
COORDONARE: Centrul de Cercetare al Universității din Pireu (Grecia)



IMPORTANȚA AI ÎN AIAS

Motor de Inteligență Artificială Adversarială
Crearea scenariilor de atac adaptate caracteristicilor organizației.

Rețele Generative Adversariale (GANs)
Detectarea atacurilor de inteligență artificială adversarială.

Inteligență Artificială Explicabilă (XAI)
Motor de recomandare pentru atenuarea suplimentară a atacurilor și pentru înțelegerea deciziilor luate de IA.

OBIECTIVE

PROTECȚIE HOLISTICĂ

Conceptualizar y desarrollar una arquitectura dConceptualizarea și dezvoltarea unei arhitecturi de servicii care integrează aplicații asistate de AI, mecanisme de decepție și tehnici de atenuare pentru protecția holistică a organizațiilor împotriva atacurilor cibernetice și a inteligenței artificiale adversariale.

SCENARII DE ATAC

Proiectarea și dezvoltarea unui motor adversarial AI inovator pentru crearea de scenarii de atac adaptate caracteristicilor infrastructurii hardware și software a organizațiilor vizate.

METODE INTELIGENTE DE DECEPȚIE INOVATOARE

Conceperea și implementarea unor metode inteligente de decepție bazate pe honeypot-uri de interacțiune ridicată, gemeni digitali (Digital Twin) și personaje virtuale (Virtual Persona).

METODE DE PROTECȚIE BAZATE PE IA

Proiectarea, dezvoltarea și evaluarea metodelor bazate pe inteligență artificială pentru detectarea și atenuarea atacurilor cibernetice, inclusiv a atacurilor de AI adversarial, precum și conceptualizarea și implementarea metodelor de colectare și fuziune a datelor.

MOTOR DE RECOMANDARE BAZAT PE XAI

Dezvoltarea și verificarea unui motor de recomandare bazat pe inteligență artificială explicabilă (XAI), care permite decizii proactive cu implicarea umană pentru a atenua complet atacurile de AI adversarial.

UTILIZARE ÎN SCENARII REALE

Evaluarea funcționalității, eficienței și eficacității platformei AIAS în scenarii reale.

METODOLOGIE

Fata 1: Identificarea cerințelor de sistem și a componentelor principale ale platformei

- Identificarea și definirea cerințelor de securitate, confidențialitate, funcționalitate și etică.
- Realizarea recenziilor SOTA (state-of-the-art) în domeniile cheie ale programului: metode de decepție, metode de detecție și atenuare bazate pe IA.
- Specificarea instrumentelor și aplicațiilor care vor fi utilizate pentru implementarea fiecărui modul AIAS.

Faza 2: Implementarea și validarea componentelor principale ale platformei
Fiecare modul va fi proiectat conform tehnicilor bine cunoscute, implementând metodele relevante și ghidurile furnizate de UE, utilizând totodată tehnologii de ultimă generație.

Fata 3: Integrare, studiu de concept și evaluare în condiții reale
Obiectivul principal al acestei faze este livrarea platformei AIAS, iar modulele care o compun trebuie să fie funcționale și să lucreze perfect împreună. După finalizarea integrării platformei, toți partenerii o vor evalua prin studii de caz pilot, selectate atent pentru scenarii reale. Această sarcină poate include modificări ale platformei pe baza feedback-ului obținut în timpul experimentelor.