FICHA DE ACCIÓN FORMATIVA

IA aplicada a la ciberseguridad

IACIBER-ADR



MODALIDAD	DURACIÓN
Online	**40 horas**

Objetivos

Aplicar técnicas de inteligencia artificial para detectar, priorizar y responder a riesgos de seguridad con supervisión humana y criterios de control, trazabilidad y cumplimiento.

Contenidos

1 - Fundamentos de ciberseguridad aplicados a la IA

Objetivo

Identificar los principios básicos de ciberseguridad (confidencialidad, integridad y disponibilidad) aplicados a entornos de inteligencia artificial.

Reconocer las amenazas y vulnerabilidades comunes en entornos digitales que pueden afectar a sistemas basados en IA.

Contenido

Fundamentos de ciberseguridad aplicados a la IA

Principios básicos de ciberseguridad aplicados a entornos de inteligencia artificial

La tríada CIA como base

Principios extendidos para entornos de IA

El ciclo de vida seguro de un sistema de IA

Amenazas y vulnerabilidades comunes en entornos digitales

1. Amenazas comunes

2. Vulnerabilidades críticas

Introducción al OWASP LLM Top 10

Explicación

Seguridad en aplicaciones que manejen modelos a gran escala

Rol de la IA en la identificación y mitigación de riesgos

1. Identificación inteligente de riesgos

Mitigación automatizada de riesgos

Limitaciones y riesgos del uso defensivo de la IA

Resumen

2 - Introducción a la inteligencia artificial en ciberseguridad

Objetivo

Describir los conceptos básicos de inteligencia artificial y aprendizaje automático relevantes para la ciberseguridad.

Explicar el papel de la IA en la prevención y detección de amenazas de seguridad en el contexto actual.

Contenido

Introducción a la inteligencia artificial en ciberseguridad

Conceptos básicos de inteligencia artificial y aprendizaje automático

1. Conceptos operativos (¿Qué hace?)

2. Aprendizaje automático (Machine Learning - ML)

El cambio de paradigma: programación tradicional vs. Machine Learning

El ciclo de vida del Machine Learning (ML Pipeline)

Conceptos críticos y desafíos técnicos

Tabla comparativa de desafíos técnicos

3. Paradigmas principales del aprendizaje automático

4. Aprendizaje profundo (Deep Learning - DL)

¿Cómo ayuda el Deep Learning a proteger sistemas y datos? (Ejemplos Concretos)

Tabla comparativa: Machine Learning Tradicional vs. Deep Learning

Contexto actual de la ciberseguridad y retos emergentes

Factores clave del escenario actual de amenazas

Retos emergentes en la era digital

1. Ciberataques potenciados por Inteligencia Artificial (IA Adversaria)

2. Ataques dirigidos a los propios sistemas de IA (AI Security)

3. Ataques a la Cadena de Suministro de Software (Supply Chain Attacks)

4. La amenaza cuántica inminente (Post-Quantum Cryptography)

Resumen de retos emergentes

La importancia de la IA para la protección de sistemas y datos

Aplicaciones de Machine Learning en Ciberseguridad

1. Detección y Prevención de Intrusiones (IDS/IPS)

2. Detección de Phishing

3. Preservación de la Privacidad

4. Detección de Spam

5. Análisis de Riesgos

6. Detección de Malware

7. Testing de Propiedades de Seguridad

Resumen técnico de aplicaciones

Resumen

3 - Aprendizaje supervisado para detección de amenazas

Objetivo

Identificar los modelos clásicos de aprendizaje supervisado aplicables a la detección de malware y ataques.

Interpretar las métricas de evaluación de modelos supervisados para valorar su eficacia en la detección de amenazas.

Contenido

Aprendizaje supervisado para detección de amenazas

Modelos clásicos de aprendizaje supervisado

Árboles de decisión y Random Forest

Comparativa árbol de decisión vs random forest

Máquinas de soporte vectorial (SVM)

Regresión logística y naive bayes

Comparativa Regresión logística y naive bayes

Aplicación práctica en detección de malware y ataques

Aplicación práctica en detección de malware

Ejemplo práctico en Python con Scikit-Learn

Ejemplo práctico en Python con Scikit-Learn

Clasificadores supervisados analizados

Análisis de resultados del modelo de detección

Introducción a los sistemas de detección de intrusos (IDS) y Machine Learning

Ejemplo práctico: clasificación multiclase de tráfico de red

Script en Python con evaluación multiclase y curva ROC-AUC

Evaluación y mejora de modelos supervisados

Métricas de evaluación

Precisión (Accuracy)

Recall (Exhaustividad o Sensibilidad)

F1-score

Curva ROC-AUC (Área Bajo la Curva ROC)

Resumen

4 - Aprendizaje no supervisado y detección de anomalías

Objetivo

Distinguir las principales técnicas de aprendizaje no supervisado y clustering utilizadas para identificar patrones anómalos en tráfico y registros.

Aplicar criterios de control y trazabilidad para validar los resultados de detección de anomalías generados por modelos no supervisados.

Contenido

Aprendizaje no supervisado y detección de anomalías

Introducción a técnicas no supervisadas y clustering

Algoritmos de particionamiento y densidad

K-Means: Particionamiento basado en centroides

DBSCAN: Detección nativa basada en densidad

Tabla comparativa: K-Means vs DBSCAN

Reducción de dimensionalidad

La "maldición de la dimensionalidad" en ciberseguridad

¿Qué es PCA y cómo simplifica los datos?

PCA aplicado a la detección de anomalías

Identificación de patrones anómalos en tráfico y registros

Detección de anomalías con K-Means en Python

Detección de anomalías con K-Means en Python (video)

Análisis de los resultados y matriz de confusión

Reporte de clasificación

PCA y Detección de Anomalías en Python

Interpretación de los resultados

Validación y control de resultados de detección de anomalías

Casos de uso y desafíos en detección temprana

Aplicaciones

Retos técnicos

Resumen

5 - Redes neuronales y deep learning en seguridad

Objetivo

Identificar las arquitecturas de redes neuronales profundas más adecuadas para la detección de intrusiones y reconocimiento de patrones maliciosos.

Analizar los resultados de un modelo de deep learning aplicado a ciberseguridad, valorando su rendimiento y los criterios de supervisión humana necesarios

Contenido

Redes neuronales y deep learning en seguridad

Fundamentos de redes neuronales profundas

Arquitectura multicapa

Componentes estructurales de una red neuronal

Extracción jerárquica y automática de características

Pesos y funciones de activación

Los pesos (Weights)

Las funciones de activación

Arquitecturas de redes neuronales: MLP vs. RNN

MLP (Perceptrón Multicapa / Multi-Layer Perceptron)

RNN (Redes Neuronales Recurrentes / Recurrent Neural Networks)

Comparativa MLP vs RNN

Detección de intrusiones y reconocimiento de patrones

Sistemas de detección de intrusiones (IDS/IPS)

Planteamiento del escenario práctico

Código de implementación en Python con Scikit-Learn

Desglose y análisis del funcionamiento en ciberseguridad

Análisis general del rendimiento

Análisis del reporte de clasificación

Conclusiones

Detección de fraude y phishing

Planteamiento del escenario práctico

Código de implementación en Python con scikit-learn

Explicación del preprocesamiento implementado:

Desglose y análisis del funcionamiento en ciberseguridad

Optimización y entrenamiento en escenarios ciberseguros

Resumen

6 - Procesamiento de lenguaje natural para análisis de amenazas

Objetivo

Aplicar técnicas de PLN para identificar indicadores de amenaza en textos de seguridad como logs, alertas y correos maliciosos.

Seleccionar herramientas y técnicas de PLN adecuadas para automatizar el análisis de registros de seguridad bajo supervisión humana.

Contenido

Procesamiento de lenguaje natural para análisis de amenazas

Introducción al procesamiento de lenguaje natural (PLN)

Aplicaciones en análisis de logs y correos maliciosos

Aplicaciones del PLN en el análisis de logs

Aplicaciones principales

Ejemplo práctico

Desarrollo en Python

Detección de correos maliciosos y phishing

Características

Ejemplo práctico

Desarrollo en Python

Desarrollo en Python (video)

Herramientas y técnicas relevantes de PLN

Resumen

7 - Evaluación de riesgos mediante IA

Objetivo

Evaluar riesgos de seguridad utilizando modelos predictivos apoyados en IA, priorizándolos según probabilidad e impacto.

Valorar los riesgos específicos de seguridad presentes en soluciones basadas en IA, identificando vectores de ataque y fallos potenciales del sistema.

Contenido

Evaluación de riesgos mediante IA

Riesgos asociados a la IA

Riesgos de reputación y cumplimiento legal y ético

Análisis de riesgos en modelos de IA

Matriz de impacto por dimensión

Matriz de evaluación de riesgos para modelos de IA

Matriz de Evaluación para "Recruiting AI"

1. Identificación y evaluación de riesgos

2. Plan de acción y mitigación

Gestión de riesgos en modelos de IA

Framework de gestión de riesgos de IA del NIST (NIST AI RMF)

Resumen

8 - Diseño de soluciones de seguridad basadas en IA

Objetivo

Diseñar una solución de seguridad basada en IA, especificando objetivo, datos de entrada, salida esperada y controles de supervisión humana.

Seleccionar arquitecturas e integrarlas con infraestructuras existentes, aplicando buenas prácticas de escalabilidad y trazabilidad.

Contenido

Diseño de soluciones de seguridad basadas en IA

Arquitecturas y componentes clave

Modelos arquitectónicos

Componentes clave del sistema

Integración con infraestructuras existentes

Conectividad con plataformas de operaciones de seguridad (SIEM, EDR y SOAR)

Convivencia con sistemas legacy e infraestructuras híbridas

Protocolos de comunicación y estandarización

Integración con el gobierno de identidades (IAM) y políticas de acceso

Herramientas para la integración de soluciones de seguridad basadas en IA

Herramientas open source (código abierto)

Herramientas de pago (comerciales y propietarias)

Buenas prácticas de diseño y escalabilidad

Garantizar la resiliencia y el control humano (HITL)

Explicabilidad y transparencia analítica (XAI)

Escalabilidad arquitectónica y gestión de picos de tráfico

Mantenimiento, ciclo de vida y seguridad de los modelos (MLOps)

Resumen

9 - Privacidad, ética y cumplimiento normativo

Objetivo

Reconocer los requisitos legales y éticos aplicables al despliegue de IA en entornos de seguridad, incluyendo el RGPD y la normativa sectorial.

Verificar el cumplimiento de criterios de privacidad y ética en una solución de IA concreta, proponiendo controles correctivos cuando sea necesario.

Contenido

- Privacidad, ética y cumplimiento normativo
- Aspectos legales y regulatorios en IA y ciberseguridad
- Reglamento de inteligencia artificial de la UE (EU AI Act)
- Directivas NIS2 y DORA (Digital Operational Resilience Act)
- NIS2
- Ámbito de aplicación y clasificación de entidades
- Medidas de gestión de riesgos de ciberseguridad
- Flujo obligatorio de notificación de incidentes
- Reglamento DORA: Resiliencia Operativa Digital para el Sector Financiero
- Los 5 pilares fundamentales de DORA
- DORA y los sistemas de inteligencia artificial
- NIS2 vs DORA
- Cyber Resilience Act (CRA)
- Seguridad desde el diseño y por defecto (Security by Design & Default)
- Gestión transparente de la cadena de suministro (AI-BOM / SBOM)
- Gestión y notificación de vulnerabilidades
- Mantenimiento y parches durante el ciclo de vida
- Principios éticos en la implementación de IA
- Equidad y mitigación del sesgo algorítmico
- Explicabilidad y transparencia (XAI)
- Supervisión y control humano (Human-in-the-Loop)
- Auditabilidad y trazabilidad en MLOps
- Protección de datos y privacidad en sistemas inteligentes
- Vectores de riesgo para la privacidad en modelos de IA
- Tecnologías de preservación de la privacidad (Privacy-Enhancing Technologies - PETs)
- Privacidad diferencial (Differential Privacy - DP)

Aprendizaje federado (Federated Learning - FL)

Cifrado homomórfico y SMPC

Matriz comparativa de técnicas de privacidad

Cumplimiento del RGPD en el ciclo de vida de la IA

El desafío del "derecho al olvido" (Machine Unlearning)

Decisiones automatizadas y elaboración de perfiles (Art. 22 del RGPD)

Evaluación de impacto en la protección de datos (EIPD / DPIA)

Resumen

10 - Herramientas y plataformas para IA en seguridad

Objetivo

Seleccionar herramientas y librerías de IA adecuadas para apoyar la detección, priorización y respuesta a riesgos de seguridad.

Configurar y desplegar una solución de IA en un entorno práctico de seguridad, verificando su correcto funcionamiento y trazabilidad.

Contenido

Herramientas y plataformas para IA en seguridad

Revisión de herramientas y librerías relevantes

Librerías de Machine Learning y Deep Learning

Frameworks de seguridad y filtrado para modelos grandes de lenguaje (LLM Guardrails)

Herramientas especializadas en ciberseguridad e inteligencia de amenazas

Plataformas de desarrollo y entornos locales

Ecosistemas de nube (MLOps y Enterprise)

Entornos locales (On-Premise)

Ollama

Instalación y ejecución

Uso de los modelos desde Python con LangChain

LM Studio

Casos de uso

Configuración y despliegue de soluciones IA

Estandarización del entorno

Infraestructura de inferencia (Model Serving)

Capa de seguridad y "Guardrails"

Observabilidad y monitoreo

Resumen

11 - Tendencias actuales y futuras en IA aplicada a la seguridad

Objetivo

Identificar las innovaciones tecnológicas emergentes en IA aplicada a la seguridad y su impacto potencial en los entornos ciberseguros

Argumentar cómo las tendencias futuras condicionan las decisiones de diseño y gestión de soluciones de seguridad basadas en IA

Contenido

Tendencias actuales y futuras en IA aplicada a la seguridad

Innovaciones tecnológicas y emergentes

Sistemas SOC autónomos y agentes inteligentes

Automatización tradicional (SOAR - Security Orchestration, Automation, and Response)

Agentes inteligentes (SOC Autónomo - Probabilístico)

Sistemas SOC autónomos vs. Automatización tradicional

Ecosistema open source para el SOC autónomo

A. Capa de datos y SIEM

B. El cerebro del agente (Orquestación y LLM)

C. Capa de acción y respuesta

Desafíos futuros y oportunidades

Nuevos vectores de ataque nativos de la IA

Es el equivalente al SQL Injection para los modelos de lenguaje, pero dirigido a la lógica probabilística del LLM.

A diferencia de un software tradicional donde las decisiones están acotadas por reglas dinámicas rígidas, un agente opera según intenciones.

Los modelos y sistemas RAG (Retrieval-Augmented Generation) dependen completamente de la calidad del conocimiento al que acceden.

La IA democratiza y automatiza los ataques de ingeniería social a gran escala con un nivel de realismo sin precedentes.

Un vector específico del modelo de costes de la IA en la nube.

La oportunidad de la respuesta hiperautomatizada

Preparación para evoluciones del entorno ciberseguro

Confianza cero adaptativa

Gobernanza e integridad de los agentes

Trazabilidad inmutable (Audit Trails) y registro de razonamiento

Visibilidad operativa y monitorización en tiempo real

Salvaguardas y mecanismos anti-manipulación (Agent Guardrails)

Resumen

12 - Ataques adversarios y robustez de modelos de IA en seguridad

Objetivo

Identificar los principales tipos de ataques adversarios dirigidos a modelos de IA en contextos de seguridad, como evasión, envenenamiento y extracción

Proponer y justificar medidas de robustez y defensa ante ataques adversarios, evaluando su viabilidad en un despliegue real con supervisión humana

Contenido

Ataques adversarios y robustez de modelos de IA en seguridad

Introducción a los ataques adversarios

Ataques de envenenamiento (poisoning)

Características de los ataques de envenenamiento

Ejemplo de ataque de envenenamiento

Panel izquierdo: datos de entrenamiento limpios (No Poisoning)

Panel derecho: datos de entrenamiento envenenados (Poisoned)

Ataques de evasión (Evasion Attacks)

Características de los ataques de evasión

Ejemplo de ataque de evasión

Técnicas de defensa ante ataques adversarios

Defenderse del aprendizaje automático adversario

MITRE ATLAS

Evaluación de robustez ante ataques adversarios

Ecosistema y Herramientas

TextAttack: El Framework de Referencia para NLP

Recetas de ataques pre-construidas (Attack Recipes)

Vía línea de comandos (CLI)

Vía API de Python

Resumen