MAJ 25/02/2025

Formation « Sécurité des systèmes d'intelligence artificielle »

Réf : SECUIA

Grâce aux avancées techniques de la dernière décennie, le déploiement fulgurant de l'IA désormais omniprésente pose de nouveaux défis de cybersécurité, avec de nombreux risques techniques peu connus et mal maîtrisés par les entreprises.

Cette formation à destination de publics techniques (développeurs, pentesters, data scientists, ...) apportera toutes les compétences nécessaires pour exploiter et remédier aux vulnérabilités les plus courantes liées aux systèmes d'intelligence artificielle, avec une attention particulière portée aux LLM et autres modèles génératifs.

Objectifs

- **Maîtriser le vocabulaire et les concepts principaux des modèles de deep learning modernes**
- **Comprendre les risques liés à la compromission d'un système d'IA par un tiers malveillant**
- **Préparer et réaliser un audit de sécurité approfondi d'un système d'IA**
- **Mettre en place des mesures correctives et des bonnes pratiques de développement pour assurer la sécurité de ces systèmes**

Durée & horaires

- **3 jours, soit 21 heures**
- **Horaires : du lundi au jeudi : de 9h30 à 12h00 et de 13h30 à 18h00**

Nombre de participants

- **Minimum 6 participants – Maximum 16 participants**

Public visé

- **Développeurs, responsables techniques, architectes travaillant sur des projets d'IA**
- **Consultants en cybersécurité souhaitant renforcer leur éventail de compétences**

Prérequis

- **Expérience en programmation, idéalement liée à l'IA ou la science des données**
- **Notions de base en sécurité applicative, la participation à la formation PENTEST1 ou PENTESTWEB est un plus**

Méthodes pédagogiques

- **Cours magistral illustré par des exemples concrets et des démonstrations**
- **Exercices de mise en pratique sur des environnements réalistes, challenges type Capture The Flag**

Supports

- **Support de cours au format papier en français en mode présentiel et au format numérique en mode distanciel**
- **Feuilles d'émargement par demi-journée de formation et certificat individuel de formation attestant de la participation à l'ensemble de la formation**

Modalité d'évaluation de la formation

- **Fiche d'évaluation remise aux stagiaires à l'issue de la formation afin de recueillir leurs impressions et identifier d'éventuels axes d'amélioration**

Certification

- **Cette formation prépare à l'examen de certification HS2 SECUIA. A l'issue de cette formation, le stagiaire passe l'examen d'une durée de 2h00 en français. L'examen est constitué d'une partie QCM et d'une partie évaluation pratique type CTF.**

Programme

Jour 1 - Vulnérabilités du modèle

- Concepts fondamentaux du deep learning : entraînement et fine-tuning, architectures, jeux de données, embeddings, ...
- Attaques antagonistes
- Injection de prompt
- Empoisonnement des données d'entraînement
- Attaques sur l'apprentissage fédéré
- Inversion de modèle

Jour 2 - Vulnérabilités applicatives

- Injection via les sorties du modèle
- Fuite de données sensibles
- Attaques sur la chaîne d'approvisionnement
- Élévation de privilèges via le modèle
- Attaques sur RAG : fuite de documents, attaques par canal auxiliaire

Jour 3 - Mesures défensives et bonnes pratiques + évaluation

- Méthodes de préparation et de restitution d'un audit de sécurité IA
- Sécurisation des systèmes d'IA, de la conception jusqu'à la mise en production
- Gestes à adopter en cas d'attaque, considérations juridiques et protection des données
- Correction des travaux pratiques des jours 1 et 2
- Examen de compétences