FORMATION : **Sécurisation de l'IA**
REFERENCE : **LMSI\SECIA**
TARIF INTER : **3200 € HT**
TARIF INTRA : **3840 € HT**

Indicateur annuel de satisfaction : 95% **Nombre de stagiaires ayant participés : 47**

OBJECTIFS DE LA FORMATION

- Comprendre les enjeux de sécurité spécifiques à l'IA
- Identifier les principales menaces et vulnérabilités
- Maîtriser les techniques de défense et bonnes pratiques
- Appliquer un cadre éthique et réglementaire

NOTRE ENGAGEMENT QUALITÉ

À QUI S'ADRESSE CETTE FORMATION

RSSI / DSI; responsables conformité et risk managers; juristes d'entreprise; dirigeants et managers opérationnels

PROFIL

RSSI / DSI; responsables conformité et risk managers; juristes d'entreprise; dirigeants et managers opérationnels

PRÉ-REQUIS

Notions de cybersécurité et compréhension générale de l'IA

LES PLUS DE CETTE FORMATION

- Formation animée par un expert reconnu en cybersécurité IA
- Approche pragmatique avec ateliers techniques (Foolbox, ART, LLM)
- Couverture complète : attaques, défenses, cadre légal AI Act
- Études de cas réels et supports techniques fournis

MODALITÉS PÉDAGOGIQUES

- Alternance théorie / pratique
- Études de cas concrets et retours d'expérience
- Ateliers en sous-groupes et travaux pratiques
- Échanges et partage entre stagiaires
- QCM d'évaluation des acquis

MOYENS DE FORMATION

Formateur expert, un ordinateur par personne avec environnement Python

MODALITÉS DE RÉALISATION

Classe à distance (prévoir une connexion adsl 8Mb/s min.) ou présentiel, la formation se déroule sur site chez le client.

MODALITÉS D'ÉVALUATION

Au cours de la formation, suivi du responsable de projet. Évaluation écrite à l'issue de la formation où le participant donnera ses impressions par écrit et à l'oral.

ORGANISATION

DURÉE

Quatre jours soit 36 heures

EFFECTIF

Minimum : 1 personne **Maximum : 6 personnes**

La session peut être annulée 48h avant la date de réalisation s'il n'y a qu'un seul participant, car pour nous la formation est aussi un moment de partage des bonnes pratiques entre les différents stagiaires.

SUIVI

Feuille d'émargement signée par ½ journée, évaluation de satis- faction et d'auto-évaluation stagiaire. Attestation de stage

INTERVENANT

Damien Lamy**,** intervient depuis plus de 20 ans dans le conseil et la formation en cybersécurité. Spécialiste de la sécurisation des systèmes d'IA, il accompagne les organisations face aux nouvelles menaces liées à l'intelligence artificielle.

RESPONSABLE PÉDAGOGIQUE ET ADMIN.

Damien LAMY 04 58 17 28 58 dlamy@lmsi.eu

CONTACT HANDICAP

Damien LAMY 04 58 17 28 58 dlamy@lmsi.eu

DATES DE FORMATION

Pour toute demande de calendrier adressez un mail formation@lmsi.eu en précisant le nom du stage.

CONTACT ET MODALITÉS D'INSCRIPTION

Pour s'inscrire à la formation ou recevoir des informations envoyer un mail à l'adresse formation@lmsi.eu en précisant le code de la formation.
Contact Lamy Damien 04 58 17 28 58

COORDONNÉES DE L'ORGANISME DE FORMATION

Lmsi sasu Capital 1000 € http://www.lmsi.eu
21 BD Maréchal de Lattre de Tassigny - 73100 Aix les Bains
(2ème étage)
Organisme de formation enregistré 84260258126
Siret: 817 999 709 00022 Code APE 6202A

Date de mise à jour 04/05/2026
Version 1.0

FORMATION : **Sécurisation de l'IA**
REFERENCE : **LMSI\SECIA**
TARIF INTER : **3200 € HT**
TARIF INTRA : **3840 € HT**

LES MODULES DE FORMATION

MODULE 1 - Introduction à la Sécurité de l'IA

Durée estimée : 3h

1.1 Pourquoi sécuriser l'IA ?

- Différences entre sécurité informatique classique et sécurité IA
- Enjeux économiques, sociaux et stratégiques
- Failles réelles : Tesla, ChatGPT, reconnaissance faciale

1.2 Panorama des systèmes IA concernés

- Machine Learning (ML)
- Deep Learning / réseaux de neurones
- LLM (Large Language Models)
- Systèmes autonomes et agents IA

1.3 Concepts fondamentaux

- Triade CIA (Confidentialité, Intégrité, Disponibilité) appliquée à l'IA
- Cycle de vie d'un modèle IA
- Surface d'attaque spécifique à l'IA

MODULE 2 - Menaces et Attaques sur les Systèmes IA

Durée estimée : 6h

2.1 Attaques sur les données

- Data Poisoning (empoisonnement des données d'entraînement)
- Manipulation et biais introduits volontairement
- Attaques sur les pipelines de données

2.2 Attaques sur les modèles

- Adversarial Attacks (boîte blanche / boîte noire)
- FGSM, PGD, Carlini & Wagner
- Model Inversion (reconstruction de données sensibles)
- Membership Inference Attack
- Model Stealing / Extraction

2.3 Attaques spécifiques aux LLM

- Prompt Injection
- Jailbreaking
- Hallucinations exploitées
- Backdoor dans les modèles pré-entraînés

2.4 Attaques sur l'infrastructure

- Compromission des APIs d'IA
- Supply chain attacks (bibliothèques, modèles tiers)
- Attaques sur les environnements cloud/GPU

MODULE 3 - Vulnérabilités et Surface d'Attaque

Durée estimée : 4h

3.1 Vulnérabilités dans le cycle de vie du modèle

- Collecte : données corrompues, non représentatives
- Entraînement : poisoning, backdoors
- Déploiement : API exposée, modèle volé
- Inférence : adversarial inputs, injection
- Maintenance : drift, mise à jour malveillante

3.2 Dépendances et supply chain

- Risques liés aux modèles open source
- Bibliothèques tierces (PyTorch, TensorFlow, HuggingFace)
- Modèles pré-entraînés non vérifiés

3.3 Facteurs humains

- Ingénierie sociale ciblant les équipes IA
- Mauvaises configurations
- Manque de sensibilisation

MODULE 4 - Techniques de Défense et Sécurisation

Durée estimée : 8h

4.1 Sécurisation des données

- Validation et nettoyage des datasets
- Détection d'anomalies dans les données
- Chiffrement et anonymisation (Differential Privacy)
- Federated Learning pour la confidentialité

4.2 Robustesse des modèles

- Adversarial Training et Certified Defenses
- Détection d'entrées adversariales
- Régularisation anti-overfitting

4.3 Sécurisation des LLM

- Filtrage des inputs/outputs et guardrails
- Red Teaming spécifique aux LLM
- Gestion des permissions et des contextes

4.4 Infrastructure et MLOps sécurisé

- Sécurisation des APIs (auth, rate limiting)
- Isolation des environnements d'entraînement
- Monitoring et détection d'anomalies en production
- CI/CD sécurisé, versioning et audit trails
- Tests de sécurité automatisés

FORMATION : **Sécurisation de l'IA**
REFERENCE : **LMSI\SECIA**
TARIF INTER : **3200 € HT**
TARIF INTRA : **3840 € HT**

MODULE 5 - Éthique, Biais et IA de Confiance

Durée estimée : 4h

5.1 Biais et discriminations

- Types de biais (données, algorithme, humain)
- Impact sociétal des biais
- Techniques de détection et correction

5.2 Explicabilité et transparence (XAI)

- Importance de l'interprétabilité pour la sécurité
- LIME, SHAP et autres outils
- Lien entre opacité et vulnérabilité

5.3 IA de confiance (Trustworthy AI)

- Principes de l'IA éthique (UE)
- Robustesse, fiabilité, sécurité
- Certifications et audits

MODULE 6 - Cadre Réglementaire et Normatif

Durée estimée : 3h

6.1 Réglementations en vigueur

- AI Act européen (2024)
- RGPD et impact sur les systèmes IA
- Recommandations de l'ANSSI
- Directives NIST AI RMF

6.2 Standards et normes

- ISO/IEC 42001 (Management de l'IA)
- ISO/IEC 27001 appliqué à l'IA
- OWASP Top 10 pour les LLM

6.3 Responsabilité et gouvernance

- Qui est responsable en cas d'incident IA ?
- Politique de sécurité IA en entreprise
- Gestion des incidents spécifiques à l'IA

MODULE 7 - Cas Pratiques et Travaux Dirigés

Durée estimée : 6h

7.1 Atelier - Attaques adversariales

- Générer des exemples adversariaux sur un modèle
- Outils : Foolbox, ART (Adversarial Robustness Toolbox)

7.2 Atelier - Prompt Injection

- Tester des techniques de jailbreaking sur un LLM
- Mettre en place des guardrails

7.3 Atelier - Audit de sécurité d'un pipeline ML

- Identifier les vulnérabilités d'un pipeline complet
- Rédiger un rapport de sécurité

7.4 Étude de cas

- Incidents réels : Microsoft Tay, failles GPT, biais Amazon RH
- Proposer des contre-mesures

MODULE 8 - Perspectives et Tendances

Durée estimée : 2h

8.1 IA générative et nouveaux risques

- Deepfakes et désinformation
- IA utilisée pour attaquer d'autres systèmes
- Automatisation des cyberattaques par l'IA

8.2 IA défensive

- IA pour la détection d'intrusions
- IA pour l'analyse de malwares
- SOC augmenté par l'IA

8.3 Recherche et avenir

- Quantum computing et IA sécurisée
- Tendances en adversarial ML
- Projets de recherche en cours (DARPA, ENISA)