

# Analyser les Menaces et Atténuer les Risques dans les Solutions d'IA

## INFORMATIONS GÉNÉRALES

**Type de formation :** Formation continue

**Éligible au CPF :** Non

**Domaine :** IA, Big Data et Bases de données

**Action collective :** Oui

**Filière :** IA

**Code ACO :** CISIA

**Rubrique :** Certification ATLAS : CISIA (Actions co.)

**Code de formation :** CISIA-MEN

## € Tarifs

**Prix public :** 3000 €

### Tarif & financement :

Financement possible via les Actions Collectives ATLAS ou le Plan de Formation.

## PRÉSENTATION

### Objectifs & compétences

**Analyser les Menaces Associées à l'Élaboration d'une Solution d'IA**  
(Compétence C2)

- Identifier les risques éthiques et sociétaux liés à l'exploitation des solutions d'IA.
- Comprendre les implications réglementaires et éthiques pour prévenir les dérives potentielles.

**Comprendre et Appliquer les Mécanismes d'Atténuation des Attaques Adversariales**  
(Compétence C4)

- Étudier les techniques et les approches pour atténuer les attaques adversariales contre les modèles d'IA.
- Mettre en œuvre des stratégies pour protéger les modèles contre ces attaques.

**Évaluer les Risques Résiduels et Assurer la Sécurité des Modèles d'IA**  
(Compétence C8)

Évaluer l'efficacité des mécanismes d'atténuation et les risques résiduels associés.  
Adapter les stratégies de sécurité en fonction des résultats de l'évaluation des risques

### Public visé

Professionnels en data science, ingénieurs en machine learning, et responsables de la sécurité des systèmes d'IA ayant une expérience préalable en développement et en déploiement de solutions d'IA.

### Pré-requis

Connaissance de base en intelligence artificielle et en machine learning  
Expérience en développement de modèles d'IA et en analyse des risques

## 📍 Lieux & Horaires

**Campus :** Ensemble des sites

**Durée :** 8 heures

**Rythme :** 9h30-12h30 et 14h-17h

### Délai d'accès :

Jusqu'à 8 jours avant le début de la formation, sous condition d'un dossier d'inscription complet

**Distanciel possible :** Oui

## 📅 Prochaines sessions

Cliquez sur la date choisie pour vous inscrire :

### ■ 17 / 07 / 2025

- 📍 : Ensemble des sites
- ✓ : Distanciel possible
- 🕒 : 8 heures
- 📅 : 1 jours

### ■ 21 / 07 / 2025

- 📍 : Ensemble des sites
- ✓ : Distanciel possible
- 🕒 : 8 heures
- 📅 : 1 jours

## PROGRAMME

### 1. Introduction aux Menaces et Risques en IA (1h)

- Présentation des objectifs pédagogiques
- Contexte et importance de la sécurité dans les solutions d'IA

### 2. Identification des Risques Éthiques et Sociétaux (2h)

- Analyse des risques éthiques et sociétaux liés aux solutions d'IA (Compétence C2)
- Étude de cas : exemples de dérives éthiques et impact réglementaire

### 3. Mécanismes d'Atténuation des Attaques Adversariales (3h)

- Théorie des attaques adversariales et des mécanismes de défense (Compétence C4)
- Application pratique : mise en œuvre de techniques de protection contre les attaques
- Cas pratiques : simulation d'attaques et application des mécanismes d'atténuation

### 4. Évaluation des Risques Résiduels et Sécurité des Modèles (1h30)

- Méthodes d'évaluation des risques résiduels (Compétence C8)
- Atelier : évaluation des impacts des mécanismes d'atténuation sur la sécurité des modèles
- Discussion sur les meilleures pratiques pour maintenir la sécurité des solutions d'IA

**5. Synthèse et Évaluation** (30 min)

- Révision des concepts clés
- Évaluation des compétences acquises à travers des exercices pratiques et un quizz

**MODALITÉS****Modalités**

L'ensemble du parcours est accessible en présentiel, à distance ou mode hybride.

**Présentation théorique** en présentiel.

**Atelier pratique** avec exercices en ligne et en présentiel.

**Études de Cas** : Analyse d'applications réelles des techniques de génération et d'augmentation.

**Discussion Interactive** : Échange sur les meilleures pratiques, les défis rencontrés et les retours d'expérience.

**CERTIFICATIONS**

A l'issue du parcours (10 modules), les candidats pourront passer le jury de certification ATLAS :

**Concevoir et implémenter une solution d'IA**