+33 (0)1 48 10 09 46
formation@ambient-it.net
140 Avenue Jean Lolive 93500 Pantin - France

Mis à jour le 25/03/2026

S'inscrire

# Formation Certification OSAI+™

ALL-IN-ONE : EXAMEN INCLUS AU TARIF AVEC LE COURS AI-300

CERTIFICATIONS OFFSEC - ACHETER VOS CERTIFICATIONS

5 jours (35 heures)

## Présentation

OSAI+ - OffSec AI Red Teamer - est une certification de référence dédiée à la sécurité offensive des systèmes d'intelligence artificielle. Positionnée à un niveau avancé, elle couvre les LLM, les architectures RAG, les agents IA, les échanges multi-agents, le Model Context Protocol, la supply chain IA et les infrastructures de déploiement modernes.

Notre formation Certification OSAI+ vous permettra de maîtriser les techniques de red teaming appliquées à l'IA dans une logique de préparation opérationnelle à la certification OffSec : reconnaissance, cartographie des surfaces d'attaque, exploitation des agents, compromission de pipelines RAG, abus des couches MCP, attaques sur la chaîne d'approvisionnement et exploitation d'infrastructures cloud et conteneurisées.

Vous apprendrez à identifier, analyser, exploiter et enchaîner les vulnérabilités propres aux systèmes IA modernes, tout en développant une méthodologie offensive rigoureuse fondée sur la compréhension des architectures, des flux de données, des relations de confiance et des mécanismes de furtivité.

À l'issue de la formation, vous serez en mesure de conduire une évaluation offensive réaliste sur un environnement IA complexe, d'exploiter des failles sur des composants LLM, RAG, agents et MCP, de modéliser les menaces spécifiques à l'IA et de produire un rapport professionnel conforme aux attentes d'un parcours de certification avancé.

Comme toutes nos formations, celle-ci s'appuie sur une approche résolument pratique, articulée autour de labs.

## Contenu de la formation

- 120 jours d'accès aux Labs en autoformation
- 10 accompagnements d'expert : 10 x demi journée (de 9h à 12h30) pour une durée totale de 35h
- 1 passage à la certification
- Accès au forum des apprenants

## Objectifs

- Comprendre les concepts fondamentaux du red teaming IA et les spécificités offensives des systèmes d'intelligence artificielle.
- Identifier et exploiter les vulnérabilités des LLM, des architectures RAG, des agents IA et des environnements MCP.
- Analyser les surfaces d'attaque d'une infrastructure IA moderne, de la couche applicative jusqu'aux composants cloud et conteneurisés.
- Mettre en œuvre des scénarios de compromission avancés incluant prompt injection, memory poisoning, data poisoning et abus de chaînes d'outils.
- Construire une stratégie de test offensive et produire un rapport technique exploitable dans un contexte de certification.
- Se préparer efficacement à la certification OSAI+ dans un cadre intensif et pratique.

## Public visé

- Pentesters
- Ingénieurs cybersécurité
- Ingénieurs IA / ML
- Architectes cloud

## Pré-requis

- Réseaux TCP/IP – Niveau intermédiaire à avancé
- Administration systèmes Linux/Windows – Niveau intermédiaire
- Scripting Bash/Python – Niveau intermédiaire
- Connaissances en cybersécurité offensive (pentest, exploitation)
- Notions de base en intelligence artificielle (LLM, APIs, RAG) recommandées
- Maîtrise de l'anglais technique

## Pré-requis logiciel

- **Kali Linux** -Téléchargeable ici

Note : Ambient IT n'est pas propriétaire de OSAI+™ cette certification appartient à OffSec ® Services LLC.

## Formation Certification OSAI+

[1ère demi-journée]

## Comprendre la sécurité des systèmes d’IA

- Introduction à la sécurité des systèmes d’IA et au positionnement de la certification OSAI+
- Différences entre red teaming traditionnel et red teaming IA
- Comprendre les enjeux de vol de données, de manipulation de modèles et de détournement de workflows
- Panorama des référentiels MITRE ATLAS et OWASP Top 10 for LLMs
- Lecture du cycle offensif : Recon, Access, Influence, Persistence, Exfiltration

[2ème demi-journée]

## Cartographier les surfaces d’attaque IA

- Identifier les surfaces d’attaque des LLM : APIs publiques, fine-tuning et déploiements locaux
- Comprendre les composants d’une architecture RAG : retrievers, bases vectorielles et pipelines d’ingestion
- Étudier les couches d’orchestration : MCP, outils et agents
- Repérer les dépendances critiques entre services IA et composants applicatifs
- Détecter les points faibles et les chemins d’attaque potentiels dans un environnement IA moderne

[3ème demi-journée]

## Reconnaissance et collecte d’informations

- Mettre en œuvre une reconnaissance passive : dépôts publics, documentation exposée, fuites et OSINT orienté IA
- Conduire une reconnaissance active : énumération d’API, fingerprinting de modèles et découverte d’endpoints
- Analyser les dépendances et cascades de confiance entre ingestion, retrieval et génération
- Comprendre les traces laissées par les systèmes IA : prompts structurés, embeddings, appels API
- Adopter des tactiques de discrétion : variation des requêtes, limitation de débit et contournement de honeypots

[4ème demi-journée]

## Attaquer les agents IA

- Comprendre l’architecture d’un agent IA : mémoire, registre d’outils, boucle de planification et messages système
- Exploiter les vulnérabilités de prompt injection directes et indirectes
- Manipuler la mémoire persistante pour influencer les décisions futures
- Exploiter les faiblesses d’invocation d’outils et les contrôles de permissions insuffisants
- Mettre en place des techniques de furtivité post-compromission dans un environnement agentique

[5ème demi-journée]

## Attaquer les systèmes multi-agents et les échanges A2A

- Étudier les architectures multi-agents et les mécanismes de coordination distribuée
- Comprendre les fondamentaux des protocoles A2A : structure des messages, transports et modèles d'authentification
- Réaliser des attaques d'impersonation, de relay et de replay sur les flux inter-agents
- Exploiter les défauts de validation, de schéma et de sérialisation
- Altérer un workflow sans le casser en conservant la continuité d'état

[6ème demi-journée]

## Exploiter les pipelines RAG

- Comprendre la chaîne RAG : ingestion, chunking, vectorisation, stockage, retrieval et synthèse LLM
- Identifier les faiblesses sur les couches de knowledge base, de retrieval et d'embeddings
- Mettre en œuvre des attaques de knowledge base leakage et de retrieval hijacking
- Exécuter des scénarios d'ingestion poisoning et d'attaque sur les embeddings
- Masquer des charges malveillantes au sein de documents légitimes pour gagner en furtivité

[7ème demi-journée]

## Exploiter le Model Context Protocol

- Comprendre le rôle du Model Context Protocol dans l'orchestration et la médiation d'outils
- Identifier les attaques de tool description poisoning et de tool shadowing
- Exploiter des scénarios de redirection d'endpoints et de contournement de contraintes
- Chaîner des outils pour obtenir une escalade de privilèges ou des actions non prévues
- Adopter des techniques de furtivité en imitant les usages légitimes et les sorties attendues

[8ème demi-journée]

## Attaquer la supply chain IA

- Cartographier la chaîne d'approvisionnement IA : datasets, poids de modèles, adaptateurs, scripts et registres
- Comprendre les mécanismes de dataset poisoning et d'insertion de triggers
- Étudier les compromissions de modèles : poids backdoorés, LoRA poisoning et couches adversariales
- Exploiter les mécanismes de distribution, de mise à jour et de confiance
- Masquer les modifications malveillantes sous une apparence conforme aux contrôles classiques

[9ème demi-journée]

## Exploiter les infrastructures IA et les déploiements

- Comprendre les architectures de déploiement IA : AWS SageMaker, Azure OpenAI, Vertex AI, serveurs self-hosted et clusters Kubernetes
- Identifier les erreurs de configuration cloud : rôles trop permissifs, endpoints publics et secrets exposés
- Exploiter les conteneurs, l'orchestration et les abus RBAC Kubernetes
- Attaquer les model servers via hot-reload, APIs de contrôle ou handlers personnalisés
- Maintenir la discrétion dans des environnements MLOps et CI/CD

[10ème demi-journée]

## Modéliser la menace et préparer l'examen

- Identifier les frontières de confiance, les dépendances et les chemins d'escalade dans un système IA
- Cartographier les actifs critiques : poids propriétaires, embeddings sensibles, endpoints MCP et sources RAG
- Construire une stratégie de Red Team IA complète et réaliste
- Structurer un rapport professionnel conforme aux attentes OffSec
- Préparer efficacement l'épreuve pratique et la phase de reporting de la certification OSAI+

## FAQ – QUESTIONS / RÉPONSES

### COMMENT SE DÉROULE L'EXAMEN POUR LA CERTIFICATION OSAI+™ ?

Vous devez absolument lire le guide officiel avant de passer votre examen.

La phase pratique de l'examen dure 24 heures et consiste à attaquer un environnement IA complet en exploitant un maximum de vulnérabilité.Après cette phase, vous aurez 24 heures pour compléter et envoyer le rapport d'exploitation où vous expliquerez votre démarche.

### DANS QUELLE LANGUE LA FORMATION OSAI+™ VOUS EST ENSEIGNÉE ?

La préparation à l'examen sera en français. Cependant, les contenus supplémentaires proposés par OffSec sont en anglais.

### L'EXAMEN POUR LA CERTIFICATION OSAI+™ EST-IL COMPRIS DANS LE PRIX DE LA FORMATION ?

Oui, vous pourrez passer l'examen après avoir suivi la formation.

### POUR QUELLE DURÉE LE LAB EST-IL ACCESSIBLE ?

Vous avez accès au lab pendant 120 jours.

### EN QUELLE LANGUE SE DÉROULE L’EXAMEN ?

L’examen se déroule en anglais.

### DOIS-JE POSSÉDER UNE WEBCAM ?

Oui, votre webcam doit être active durant la totalité de votre examen, elle doit pouvoir filmer toute votre pièce.

### DOIS-JE AVOIR UNE BONNE CONNEXION INTERNET ?

Oui, car votre ordinateur doit supporter pendant 24h un flux TeamViewer tout en attaquant en permanence des machines.

### QUEL EST LE PRIX DE LA CERTIFICATION ?

La certification seule est à 1749 euros.

## Sociétés concernées

Cette formation s’adresse à la fois aux particuliers ainsi qu’aux entreprises, petites ou grandes, souhaitant former ses équipes à une nouvelle technologie informatique avancée ou bien à acquérir des connaissances métiers spécifiques ou des méthodes modernes.

## Positionnement à l'entrée en formation

Le positionnement à l'entrée en formation respecte les critères qualité Qualiopi. Dès son inscription définitive, l'apprenant reçoit un questionnaire d’auto-évaluation nous permettant d’apprécier son niveau estimé sur différents types de technologies, ses attentes et objectifs personnels quant à la formation à venir, dans les limites imposées par le format sélectionné. Ce questionnaire nous permet également d’anticiper certaines difficultés de connexion ou de sécurité interne en entreprise (intraentreprise ou classe virtuelle) qui pourraient être problématiques pour le suivi et le bon déroulement de la session de formation.

## Méthodes pédagogiques

Stage Pratique : 60% Pratique, 40% Théorie. Support de la formation distribué au format numérique à tous les participants.

## Organisation

Le cours alterne les apports théoriques du formateur soutenus par des exemples et des séances de réflexions, et de travail en groupe.

## Validation

À la fin de la session, un questionnaire à choix multiples permet de vérifier l’acquisition correcte des compétences.

## Sanction

Une attestation sera remise à chaque stagiaire qui aura suivi la totalité de la formation.


Organisme de formation enregistré sous le numéro 11 75 54743 75. Cet enregistrement ne vaut pas agrément de l’État.
© Ambient IT 2015-2026. Tous droits réservés. Paris, France - Suisse - Belgique - Luxembourg