NYXORA

Sécuriser l'IA générative et les agents IA

Formation Nyxora · 1 jour (7 heures) · Présentiel ou visio en direct

> *Du prototype à la production : sécuriser les applications LLM et les agents IA que vos équipes construisent, achètent ou utilisent.*

Pourquoi cette formation

Les applications LLM et les agents IA ouvrent une nouvelle surface d'attaque. Un simple document, un email ou une page web peut contenir des instructions cachées qui détournent un assistant. Un agent connecté à vos outils peut envoyer des données ou exécuter des actions à la place d'un attaquant.

Ces risques ne se traitent pas avec les outils de sécurité classiques seulement. En une journée intensive, cette formation donne aux équipes les réflexes et les méthodes pour concevoir, tester et exploiter des systèmes IA sûrs, sans freiner les projets.

Elle est animée par des praticiens qui conçoivent et mettent en production des agents IA au quotidien.

Objectifs pédagogiques

À la fin de la formation, le participant est capable de :

1. **Identifier** les principales menaces propres aux applications LLM et aux agents IA, en s'appuyant sur les référentiels OWASP (Top 10 LLM 2025 et Top 10 Agentic Applications 2026).
2. **Expliquer** le fonctionnement des attaques majeures : prompt injection directe et indirecte, fuite de données, détournement d'agent, attaques sur la supply chain et les serveurs MCP.
3. **Reproduire** les attaques majeures sur des plateformes d'entraînement et proposer les défenses adaptées.
4. **Concevoir** une défense en profondeur : moindre privilège, validation humaine, isolation, filtrage, journalisation.
5. **Situer** les obligations applicables : AI Act, RGPD et lois nationales sur les données (France, Europe, Maghreb).
6. **Construire** un plan d'action adapté à son organisation.

Public

- Développeurs et ingénieurs IA, data ou logiciel.
- Architectes, DevOps et équipes produit.
- RSSI, équipes cybersécurité, DSI.
- Chefs de projet IA et responsables de l'innovation.

Groupe de 6 à 12 participants pour garantir le temps de pratique.

Prérequis

- Connaître le principe d'un LLM et avoir déjà utilisé un assistant IA (ChatGPT, Claude, Copilot ou équivalent).
- Créer un compte gratuit sur PortSwigger Web Security Academy avant la formation (lien dans la convocation).
- Aucune compétence en programmation n'est obligatoire. Les labs sont guidés et proposent des niveaux avancés pour les profils techniques.
- Un quiz de positionnement est envoyé avant la formation.

Programme

- **Module 1** — Pourquoi la sécurité de l'IA est différente : surface d'attaque, incidents réels, référentiels.
- **Module 2** — Prompt injection et jailbreak : attaques directes et indirectes, fuite du prompt système.
- **Module 3** — Données et sorties : fuite d'informations, shadow AI, faiblesses des RAG, sorties dangereuses, hallucinations, coûts non maîtrisés.
- **Module 4** — Sécurité des agents IA et des serveurs MCP : détournement d'objectif, abus d'outils et de privilèges, mémoire, supply chain.
- **Atelier pratique (1 h 30)** — Labs en ligne sur des applications intégrant un LLM (Gandalf, PortSwigger Web Security Academy), puis défense de l'agent d'Atlas Services avec la checklist sécurité.
- **Module 5** — Défense en profondeur et red teaming.
- **Module 6** — Gouvernance, conformité et réponse à incident : AI Act, RGPD et lois locales, deepfakes, ISO/IEC 42001.
- **Plan d'action** individuel.

Méthodes pédagogiques

- Un atelier pratique de 1 h 30 sur deux plateformes d'entraînement en ligne, gratuites et reconnues, suivi d'un exercice de défense sur le cas Atlas Services.
- Démonstrations en direct et études de cas.
- Rien à installer : un navigateur suffit, en salle comme en visio.
- Cas réels issus des projets du studio (anonymisés).
- Les participants signent une charte d'usage éthique : les techniques apprises s'utilisent uniquement sur des systèmes autorisés.

Ce que reçoit chaque participant

- **Guide apprenant** : le cours approfondi, les slides et les fiches mémo (PDF), le cahier participant (énoncé de l'atelier, quiz, plan d'action) et la charte à signer.
- **Kit de templates** :
  - politique d'usage de l'IA générative ;
  - checklist de sécurité d'un agent IA ;
  - grille d'audit d'une application LLM ;
  - grille de red teaming ;

  - procédure de réponse à incident IA.
- Fiches mémo : OWASP Top 10 LLM, OWASP Top 10 Agentic, calendrier AI Act (dans le guide apprenant).
- Les plateformes d'entraînement restent accessibles gratuitement après la formation.
- **Suivi** : 4 séances de questions-réponses en visio, une par semaine pendant le mois qui suit la formation.
- Attestation de fin de formation.

Évaluation

- **Avant** : quiz de positionnement.
- **Pendant** : labs réalisés et livrable de l'exercice de défense.
- **Après** : quiz final (mêmes notions que le quiz de positionnement) et plan d'action individuel.
- Questionnaire de satisfaction à chaud, puis à froid après la dernière séance de suivi.

Formats et modalités

Format	Durée	Lieu
Intra-entreprise	1 jour	Dans vos locaux ou en visio
Inter-entreprises	1 jour	Paris, Tunis ou en visio

- **Langues** : français ou anglais.
- **Adaptation sectorielle** en intra : cas pratiques adaptés à votre secteur et à vos outils.
- **Parcours combiné** : avec la formation "ISO/IEC 42001 en pratique et AI Act" (1 jour).
- **Délai d'accès** : selon le calendrier des sessions, nous consulter.
- **Tarif** : sur devis, selon le format (intra ou inter-entreprises).

Accessibilité

Formation accessible aux personnes en situation de handicap. Contactez-nous avant l'inscription pour étudier les adaptations nécessaires : contact@nyxora.studio.

Contact

Nyxora — Studio Data & IA · Europe & Maghreb contact@nyxora.studio · nyxora.studio