

Mis à jour le 01/10/2026

S'inscrire

Formation OWASP GenAI LLM (Top 10)

2 jours (14 heures)

Présentation

Le référentiel OWASP Top 10 for LLM Applications identifie les principaux risques de sécurité des applications d'intelligence artificielle générative. Avec les injections de prompt, la divulgation de données, les vulnérabilités du RAG et les permissions excessives des agents IA, il fournit un cadre pour analyser les menaces et structurer les protections applicatives. (genai.owasp.org)

Notre formation OWASP GenAI LLM (Top 10) vous permettra de maîtriser les dix risques officiels et de les appliquer aux architectures rencontrées lors du développement et de la mise en production.

Vous apprendrez à cartographier les surfaces d'attaque, à protéger les données et les bases vectorielles, à contrôler les sorties des modèles et à encadrer les appels d'outils selon le moindre privilège.

À l'issue de la formation, vous serez en mesure d'évaluer une application LLM, de prioriser ses vulnérabilités et de mettre en œuvre une défense en profondeur adaptée aux impacts métier.

Cette formation aborde également les tests adversariaux, la surveillance, les budgets de consommation et l'intégration des contrôles dans une démarche DevSecOps, avec des ateliers sur une application pédagogique isolée.

Comme toutes nos formations, celle-ci vous présentera **la dernière version stable** de la technologie et ses nouveautés.

Objectifs

- Identifier et contextualiser les dix risques OWASP dans une application d'intelligence artificielle générative
- Construire un modèle de menace pour une architecture LLM, RAG ou agentique
- Mettre en œuvre des protections contre les injections de prompt et les divulgations de données
- Sécuriser les pipelines RAG, les sorties des modèles et les appels d'outils
- Évaluer les protections par des tests adversariaux et prioriser les corrections
- Intégrer la surveillance et les contrôles de sécurité dans une démarche DevSecOps

Public visé

- Développeurs d'applications IA
- Architectes logiciels et IA
- Ingénieurs DevSecOps
- Ingénieurs AppSec
- Analystes sécurité

Pré-requis

- Expérience du développement applicatif et lecture de code en Python
- Connaissance des échanges HTTP, des API et du format JSON
- Compréhension des principes d'authentification, d'autorisation et de gestion des secrets
- Notions sur les LLM, les prompts et le fonctionnement du RAG

Pré-requis techniques

- Ordinateur avec processeur quatre cœurs, 16 Go de mémoire vive et 20 Go d'espace disque disponible
- Système Windows 11 avec WSL 2, macOS ou Linux, avec droits d'installation des outils pédagogiques
- Environnement Docker avec Docker Compose opérationnel pour isoler les applications des ateliers
- Installation de Python 3.12, Git et d'un éditeur tel que Visual Studio Code
- Connexion Internet stable autorisant les services du laboratoire, avec compte d'accès à un modèle hébergé à préparer selon les consignes pédagogiques, sans carte graphique dédiée requise

Programme de notre formation OWASP GenAI LLM (Top 10)

[Jour 1 - Matin]

Surfaces d'attaque et protection du contexte

- Lecture du Top 10 OWASP édition 2026 et repérage des composants d'une application LLM, d'un pipeline RAG et d'un agent IA
- Construction d'un modèle de menace : flux de données, frontières de confiance, actifs sensibles et impacts métier

- LLM01:2026 Prompt Injection : injections directes, indirectes, multimodales et persistantes via documents, RAG, mémoire, outils et contenus externes
- LLM02:2026 Sensitive Information Disclosure : minimisation des données, protection des secrets et filtrage des journaux et des traces
- LLM08:2026 Hidden Context Exposure : exposition du prompt système et du contexte interne, séparation des secrets et des décisions d'autorisation
- Atelier pratique : cartographier un assistant pédagogique, tester des injections sur des données fictives et comparer les résultats avant et après ajout de protections.

[Jour 1 - Après-midi]

Sécurisation des données et du RAG

- LLM04:2026 Supply Chain : modèles, datasets, dépendances, LoRA/PEFT et pipelines de conversion ; provenance, signatures, hashes et inventaires SBOM/AIBOM
- LLM05:2026 Data and Model Poisoning : empoisonnement des corpus, données d'entraînement et modèles, validation des sources et suivi des modifications
- LLM09:2026 Vector and Embedding Weaknesses : accès non autorisés, risques liés aux embeddings et cloisonnement des bases vectorielles
- Protection de l'ingestion documentaire : classification, traçabilité, contrôle d'intégrité et traitement des contenus non fiables
- Contrôles d'accès avant la recherche documentaire : autorisations par document, isolation des utilisateurs et séparation des caches
- Atelier pratique : introduire un document malveillant dans un corpus de test, reproduire une fuite entre utilisateurs puis vérifier les contrôles d'ingestion et d'autorisation.

[Jour 2 - Matin]

Contrôle des agents et des sorties

- LLM03:2026 Excessive Agency : fonctionnalités, permissions et autonomie excessives, application du moindre privilège aux outils
- LLM10:2026 Improper Output Handling : sorties non fiables, risques d'injection dans les systèmes aval et validation avant utilisation
- Validation des sorties structurées : schémas, contrôles métier, encodage selon le contexte et absence d'exécution implicite
- Validation du code généré avant compilation, exécution ou intégration CI/CD
- Encadrement des appels d'outils : listes d'opérations autorisées, validation des arguments et autorisation déterministe côté application
- Confinement des agents IA : isolation d'exécution, validation humaine des actions sensibles et complémentarité avec le référentiel agentique OWASP
- Atelier pratique : tester un agent doté d'un outil simulé trop permissif, restreindre ses opérations et vérifier le blocage des actions non autorisées et des sorties dangereuses.

[Jour 2 - Après-midi]

Fiabilité, maîtrise des coûts et validation sécurité

- LLM07:2026 Misinformation : hallucinations, réponses trompeuses, vérification des sources et validation humaine selon les impacts métier
- LLM06:2026 Unbounded Consumption : épuisement des ressources, dérive des coûts, modèles de raisonnement, boucles agentiques et extraction de modèle ; quotas et limites d'exécution
- Construction d'une campagne de tests adversariaux : scénarios associés aux dix risques, critères de réussite et tests de non-régression
- Intégration DevSecOps : contrôles automatisés, seuils de validation et revue des changements de modèles, de prompts et de corpus
- Organisation de la surveillance sécurité : traces expurgées, alertes, arrêt contrôlé et plan de remédiation priorisé
- Atelier pratique : exécuter une campagne de validation sur l'application fil rouge, tester les limites de consommation et produire une matrice des dix risques avec preuves, protections et risques résiduels.

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.

[Page Web du Programme de Formation](#) - Annexe 1 - Fiche 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