

Gouvernance et sécurité opérationnelle de l'IA Générative

2 jours (14 heures)

Initiation

Présentiel ou classe à distance

à partir de

1750 € HT par personne (TVA en vigueur) Repas inclus en présentiel

Contenu mis à jour le 25/09/2026. Document téléchargé le 07/10/2026.

Objectifs

A l'issue de cette formation, vous serez capable de :

- Identifier les enjeux de gouvernance spécifiques à l'IA Générative et le positionnement de cette technologie dans le SI de l'entreprise
- Décrire le cadre réglementaire et normatif actuel (EU AI Act, RGPD, NIS2 et normes de management de l'Intelligence Artificielle (IA)) appliqué aux systèmes GenAI
- Identifier, cartographier et évaluer les risques opérationnels, techniques, juridiques et organisationnels liés aux usages de l'IA Générative
- Définir une organisation de gouvernance IA (rôles, comités, processus) et des politiques d'usage encadrant les pratiques des utilisateurs
- Mettre en pratique les réflexes de sécurité opérationnelle pour les LLM / RAG / agents (vulnérabilités typiques, attaques, mesures de protection et de monitoring)
- Concevoir un plan d'action concret pour sécuriser et mettre en conformité un ou plusieurs cas d'usage d'IA Générative en production.

Les compétences visées par cette formation font l'objet d'une évaluation.

À qui s'adresse cette formation ?

Chefs de projets IT, Product Owners, responsables innovation ou digitaux, responsables conformité, Risk Managers, DPO et juristes IT, ingénieurs et architectes système / Cloud impliqués dans des projets d'Intelligence Artificielle (IA) ou GenAI, ingénieurs sécurité, analystes SOC, RSSI et responsables Cybersécurité, Data Engineers, MLOps / LLMOps Engineers et équipes de plateforme IA.

Prérequis

Avoir une bonne compréhension du fonctionnement général d'un SI et des connaissances de base en Cybersécurité (principales menaces, mesures de protection...). Avoir une familiarité avec les concepts de Data Management (qualité des données, gouvernance des données, RGPD) et du cycle de vie des applications ainsi qu'une expérience pratique dans un rôle technique ou de pilotage de projets IT. Une première exposition, même informelle, aux outils d'IA Générative (ChatGPT, Claude, Copilot...) est un plus.

Programme

1 Pendant votre formation

Gouvernance et sécurité opérationnelle de l'IA Générative

Jour 1 – Matin

Taxonomie et fondamentaux de l'IA Générative

- Architectures des modèles génératifs :
 - LLM
 - Diffusion
 - Multimodal
- Fonctionnement d'un LLM :
 - Tokenization
 - Inférence
 - Contexte
 - Température
- Panorama des modèles récents :
 - GPT
 - Claude
 - Gemini
 - Llama
 - Mistral
- Cas d'usage entreprise :
 - Génération de code
 - RAG
 - Agents autonomes
 - Assistants conversationnels

- Différence entre fine-tuning, prompt engineering et RAG

Cadre réglementaire européen et international

- Les 9 principes de gouvernance de l'Intelligence Artificielle (IA)
- Mise en place d'un comité de gouvernance IA cross-fonctionnel
- Rôles et responsabilités
- Cycle de vie d'un système IA
- Documentation :
 - Model Cards
 - System Cards
 - Registre des traitements IA
- Evaluation de la maturité IA de l'organisation

Exemples de travaux pratiques (à titre indicatif)

- Cartographier les systèmes IA d'une entreprise proposée par les stagiaires ou le formateur
- Attribuer les niveaux de risque EU AI Act
- Définir la structure de gouvernance associée :
 - Comité
 - Rôles
 - Processus de validation

Jour 1 – Après-midi

Shadow AI : détection et maîtrise

- Définition, périmètre et statistiques du Shadow AI
- Risques associés :
 - Fuite de données
 - Non-conformité
 - Absence d'audit trail
- Framework en 5 piliers :
 - Accept
 - Enable
 - Assess
 - Restrict
 - Eliminate
- Outils de détection et de monitoring du Shadow AI
- Politique d'usage acceptable de l'IA et programme de sensibilisation

Exemples de travaux pratiques (à titre indicatif)

- Réaliser un audit Shadow AI sur un scénario d'entreprise :

- *Inventaire des usages non sanctionnés*
- *Classification des risques*
- *Rédaction d'une politique d'usage acceptable*

Gestion des risques de l'IA Générative

- Taxonomie des risques appliqués à l'IA Générative :
 - Biais
 - Hallucinations
 - Fuite de données
 - Manipulation
 - Sur-autonomie des agents NIST AI RMF (Risk Management Framework)
- Analyse d'impact sur les droits fondamentaux (FRIA)
- Risques spécifiques aux systèmes RAG :
 - Empoisonnement de données
 - Manipulation de contexte...
- Risques des agents IA autonomes :
 - Excessive Agency
 - Escalade de privilèges...
- Risques liés à la Supply Chain IA :
 - Modèles tiers
 - Dépendances
 - Intégrité des poids...
- Matrice de risques et plan de traitement

Exemples de travaux pratiques (à titre indicatif)

- *Conduire une analyse de risques complète sur un cas d'usage d'IA Générative (chatbot RAG interne) en appliquant le framework NIST AI RMF :*
 - *Identification des risques*
 - *Evaluation*
 - *Mesures de mitigation*

Jour 2 – Matin

OWASP Top 10 LLM : vulnérabilités et contre-mesures

- LLM01 : Prompt Injection (directe et indirecte)
- LLM02 : Sensitive Information Disclosure
- LLM03 : Supply Chain Vulnerabilities
- LLM04 : Data and Model Poisoning
- LLM05 : Improper Output Handling

- LLM06 : Excessive Agency
- LLM07 : System Prompt Leakage
- LLM08 : Vector and Embedding Weaknesses (vulnérabilités RAG)
- LLM09 : Misinformation (hallucinations)
- LLM10 : Unbounded Consumption
- Stratégies de défense pour chaque vulnérabilité

Exemples de travaux pratiques (à titre indicatif)

- *Exploiter et défendre un LLM vulnérable :*
 - *Réaliser des attaques de Prompt Injection directe / indirecte*
 - *Tenter une extraction de System Prompt*
 - *Implémenter les contre-mesures*

Red Teaming et tests de sécurité de l'IA Générative

- Méthodologie de Red Teaming appliquée aux systèmes GenAI
- Threat Modeling spécifique LLM / RAG / Agents
- Techniques d'attaque avancées :
 - Jailbreaking multi-étapes
 - Injection indirecte via contexte
 - Escalade de privilèges d'agents
- Outils de Red Teaming automatisé :
 - Garak
 - Promptfoo
 - DeepTeam...
- Scénarios de test :
 - Exfiltration de données
 - Contournement de guardrails
 - Empoisonnement RAG
- Rapport de Red Team et plan de remédiation

Exemples de travaux pratiques (à titre indicatif)

- *Conduire une campagne de Red Teaming sur une application RAG de test :*
 - *Threat Modeling*
 - *Exécution de scénarios d'attaque (jailbreak, injection indirecte, empoisonnement de contexte)*
 - *Rédaction du rapport de findings*

Jour 2 – Après-midi

Guardrails, LLMOps et sécurisation du pipeline IA

- Architecture de sécurité d'un pipeline LLMOps : du développement à la production

- Guardrails input / output :
 - Détection d'injections
 - Filtrage PII
 - Détection de toxicité
 - Validation de format
- Frameworks de guardrails :
 - Guardrails AI
 - NVIDIA NeMo Guardrails
 - LlamaGuard...
- Prompt versioning, évaluation et monitoring en production
- Observabilité IA :
 - Monitoring de drift
 - Détection d'hallucinations
 - Métriques de qualité
- Contrôle d'accès aux modèles, audit logs et PII redaction
- Intégration de la sécurité dans le pipeline CI/CD ML (MLSecOps)
- Session tainting, circuit breakers et dual-layer permissions pour les agents

Conformité opérationnelle et amélioration continue

- Mise en oeuvre d'un AIMS conforme ISO/IEC 42001
- Documentation technique exigée par l'EU AI Act :
 - Gestion des risques
 - Gouvernance des données
 - Journalisation
 - Supervision humaine
- Post-market monitoring et signalement d'incidents
- Audit interne et évaluation de conformité
- Veille réglementaire continue et adaptation des processus
- Indicateurs de performance de la gouvernance IA (KPI)
- Feuille de route d'implémentation

Exemples de travaux pratiques (à titre indicatif)

- *Elaborer un plan de conformité EU AI Act pour un système IA à haut risque en rédigeant la documentation technique :*
 - *Fiche système*
 - *Evaluation des risques*
 - *Mesures de supervision humaine*
- *Définir les KPI de gouvernance et le calendrier d'audit*

Financement

Plusieurs dispositifs de financement sont accessibles via les OPCO (Opérateurs de Compétences), organismes agréés par le ministère du Travail dont le rôle est d'accompagner, collecter et gérer les contributions des entreprises au titre du financement de la formation professionnelle.

Pour plus d'information, une équipe de gestionnaires SKOLAE Formation spécialisée vous accompagne dans le choix de vos formations et la gestion administrative.

[Nous contacter pour en savoir plus →](#)

Modalités pédagogiques

Formation délivrée en présentiel ou distanciel* (blended-learning, e-learning, classe virtuelle, présentiel à distance).

Le formateur alterne entre méthode** démonstrative, interrogative et active (via des travaux pratiques et/ou des mises en situation).

Variables suivant les formations, les moyens pédagogiques mis en oeuvre sont :

- Ordinateurs Mac ou PC (sauf pour certains cours de l'offre Management), connexion internet fibre, tableau blanc ou paperboard, vidéoprojecteur ou écran tactile interactif (pour le distanciel)
- Environnements de formation installés sur les postes de travail ou en ligne
- Supports de cours et exercices

En cas de formation intra sur site externe à M2i, le client s'assure et s'engage également à avoir toutes les ressources matérielles pédagogiques nécessaires (équipements informatiques...) au bon déroulement de l'action de formation visée conformément aux prérequis indiqués dans le programme de formation communiqué.

* nous consulter pour la faisabilité en distanciel

** ratio variable selon le cours suivi

[Accessibilité aux personnes en situation de handicap](#)

Détails de la session

Du 17 décembre 2026 au 18 décembre 2026 (2 jours / 14 heures)

Mulhouse

Du 17 décembre 2026 au 18 décembre 2026

Mulhouse

- 17 décembre 2026 9h00 – 16h30
- 18 décembre 2026 9h00 – 16h30