Intelligence artificielle : souveraineté et sécurité

Prix : 950 €HT

Durée : 1 jour

Code de Référence : IASS

Catalogue Intelligence Artificielle

L'intelligence artificielle générative s'est imposée en quelques mois comme un outil de production massivement adopté dans les organisations publiques et privées. Cette adoption rapide déplace deux lignes de risque jusqu'ici traitées séparément : la souveraineté numérique, où partent les données et les prompts, qui contrôle les modèles et la puissance de calcul, peut-on réversibiliser un usage ; et la sécurité, une surface d'attaque inédite (injection de prompt, fuite de données, empoisonnement de modèle, agentivité excessive) qui échappe aux défenses traditionnelles.

Le cadre se durcit en parallèle : lois extraterritoriales américaines (Cloud Act, FISA), exigences du RGPD appliquées aux traitements d'IA et entrée en application progressive du règlement européen sur l'IA (AI Act). Choisir, sécuriser et piloter ses usages d'IA en maîtrisant ses données et ses dépendances est devenu incontournable.

En une journée, cette formation donne aux participants les clés pour comprendre les risques juridiques, géopolitiques et de sécurité liés à l'IA, identifier les solutions souveraines et de confiance disponibles, et poser les bases d'une stratégie d'IA maîtrisée, sécurisée et durable. La partie théorique s'appuie sur des cas génériques transposables (le « quoi faire » et le « comment faire ») et sur des retours d'expérience d'exploitation de failles réelles, suivis de l'analyse de ce qu'il aurait fallu mettre en place pour les éviter.

Objectifs de la formation

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

- Identifier les enjeux de souveraineté sur l'ensemble de la chaîne de valeur de l'IA : données, modèles, infrastructure, couche d'inférence et applicatif.
- Analyser les impacts des législations extra-européennes (Cloud Act, FISA) et du cadre réglementaire (RGPD, AI Act, normes sectorielles) sur les usages d'IA de son organisation.
- Caractériser la surface d'attaque spécifique aux systèmes d'IA et les principales failles, à l'aide d'un référentiel reconnu (OWASP Top 10 for LLM Applications).
- Distinguer les principaux acteurs, modèles et labels de l'IA souveraine et de confiance en France et en Europe.
- Sélectionner et sécuriser une solution d'IA compatible avec les exigences de souveraineté et de conformité.
- Esquisser une stratégie de gouvernance et de sécurité de l'IA adaptée à son organisation.

Public

Cette formation s'adresse aux décideurs IT, RSSI et responsables sécurité, DPO et responsables conformité, Architectes Cloud / Data / IA, responsables de projets numériques et référents IA, juristes IT et consultants en transformation digitale.

Prérequis

Aucun prérequis

Programme de la formation

Les démonstrations, exemples et travaux pratiques sont donnés à titre indicatif et pourront être adaptés selon les besoins et le contexte de la session.

Jour 1
Matinée – Fondements de la souveraineté de l'IA, cadres réglementaires et menaces
Chapitre 1 – Définir la souveraineté de l'intelligence artificielle

- Définition et dimensions de la souveraineté de l'IA : juridique, technique, données et computationnelle (accès à la puissance de calcul / GPU)
- La chaîne de valeur de l'IA et ses points de dépendance : données d'entraînement, modèles de fondation, hébergement et infrastructure, couche d'inférence (API), applicatif
- Notions clés : localisation des données et des modèles, portabilité et réversibilité, indépendance vis-à-vis d'un fournisseur (lock-in), maîtrise des traitements
- Typologies de déploiement de l'IA : API SaaS d'éditeurs, modèles managés en Cloud, IA hébergée en Cloud souverain, IA auto-hébergée (on-premise), déploiement déconnecté (air-gap) ; modèles propriétaires vs modèles ouverts (open weights)
- IA souveraine vs IA de confiance : distinctions et enjeux (maîtrise juridique et technique complète vs garanties contractuelles et labels d'un acteur pouvant rester soumis à un droit étranger)

Exemples et travaux pratiques :

- Étude de cas : exemples d'atteintes à la souveraineté via l'IA (fuite de données sensibles via des prompts vers une API extra-européenne ; dépendance critique à un modèle propriétaire modifié, restreint ou retiré)
- Identification des enjeux de maîtrise des données, des modèles, des accès et des infrastructures

Chapitre 2 – Cadres réglementaires et normatifs applicables à l'IA

- **RGPD appliqué à l'IA** : base légale des traitements, données personnelles dans les prompts et les corpus RAG, données d'entraînement, transferts hors UE, droit à l'effacement face à un modèle déjà entraîné, minimisation
- **AI Act (Règlement UE 2024/1689)** : approche par les risques (pratiques interdites / haut risque / risque limité / minimal), obligations des modèles à usage général (GPAI), exigences de transparence. Calendrier d'application et impact du Digital Omnibus
- **Réglementations sectorielles et cadres de référence** : HDS (santé), DORA (finance), NIS2, ISO/IEC 42001 (système de management de l'IA)

- **SecNumCloud et hébergement de l'IA** : objectifs, portée, acteurs qualifiés ; recommandations de l'ANSSI sur l'IA générative ; articulation avec la doctrine « Cloud au centre »
- Cas concrets de non-conformité et sanctions (jusqu'à 35 M€ ou 7 % du chiffre d'affaires mondial pour l'AI Act ; sanctions RGPD)

Exemples et travaux pratiques :

- Identification des types de données sensibles à protéger selon le RGPD dans un pipeline d'IA (prompts, RAG, fine-tuning)
- Analyse de l'impact d'une contrainte réglementaire (ex. donnée de santé soumise à HDS) sur le choix d'une solution d'IA

Chapitre 3 – Surface d'attaque, failles et risques géopolitiques de l'IA

- **Lois extraterritoriales (Cloud Act, FISA, Patriot Act)** : portée et implications pratiques pour les fournisseurs d'IA non européens, accès potentiel aux données traitées (prompts, documents RAG), y compris hébergées en région UE
- **La surface d'attaque spécifique aux systèmes d'IA** : panorama structuré à partir du référentiel OWASP Top 10 for LLM Applications (2025)
    - Injection de prompt directe et indirecte ; divulgation d'informations sensibles ; risques de la chaîne d'approvisionnement (modèles, datasets, plugins)
    - Empoisonnement des données et du modèle ; mauvais traitement des sorties ; agentivité excessive
    - Fuite de prompt système ; faiblesses des vecteurs et embeddings (RAG) ; désinformation / hallucinations ; consommation non maîtrisée
- **Dépendance technologique et risques géopolitiques** : concentration sur quelques fournisseurs de modèles de fondation, dépendance matérielle (GPU), évolution unilatérale des conditions d'usage, risque de retrait ou de censure d'un modèle

Exemples et travaux pratiques :

- Analyse critique des clauses à risque d'un contrat fournisseur d'IA type (réutilisation des données pour l'entraînement, localisation, sous-traitants, juridiction applicable)
- Étude d'un cas concret d'exploitation de faille (REX – ex. EchoLeak / injection de prompt indirecte) et identification des points de vigilance et des mesures qui auraient dû être en place

Après-midi – Solutions souveraines, sécurisation, gouvernance et stratégie
Chapitre 4 – Identifier et qualifier les solutions d'IA souveraine

- **Éditeurs et modèles européens** : Mistral AI (API et modèles ouverts) ; modèles open-weights performants (Llama, Qwen, etc.) avec analyse de licence et d'origine
- **Infrastructures souveraines françaises et européennes** : OVHcloud, Scaleway, NumSpot, Outscale (offres GPU / inférence, qualifications SecNumCloud)
- **Stack d'IA auto-hébergée** : moteurs d'inférence (Ollama, vLLM), interfaces (Open WebUI, LibreChat), base vectorielle (Qdrant), supervision (Prometheus / Grafana). Point de vigilance : les modèles « cloud-tagged » qui font sortir les données de l'infrastructure maîtrisée
- **Offres des hyperscalers à dimension souveraine** : AWS European Sovereign Cloud, Microsoft Cloud for Sovereignty, services d'inférence en régions UE et leurs limites au regard du Cloud Act

- **Labels et certifications** : analyse comparative de SecNumCloud, HDS, ISO 27001, ISO/IEC 42001 et EUCS
- Place de l'open source et des modèles ouverts dans une démarche de souveraineté : atouts (réversibilité, audit, air-gap) et limites (responsabilité, support, sécurité de la chaîne d'approvisionnement)

Exemples et travaux pratiques :

- Construction d'une grille d'analyse pour auditer une solution d'IA sur les critères de souveraineté et de sécurité
- Comparaison de plusieurs scénarios (API SaaS extra-UE vs API souveraine vs solution auto-hébergée) selon des critères juridiques, techniques, opérationnels et contractuels

Chapitre 5 – Sécuriser et gouverner une solution d'IA

- **Classification des données dans les usages d'IA** : ce qui peut ou ne peut pas être transmis à un modèle, selon la sensibilité, la criticité et les exigences réglementaires (prompts, RAG, fine-tuning)
- **Mesures de sécurité techniques** : garde-fous (guardrails) en entrée et en sortie, filtrage et prévention de fuite de données (DLP), cloisonnement des contextes, chiffrement, gestion des secrets et des clés, RBAC et étiquetage des sources RAG
- **Sécurité de l'inférence et de la chaîne agentique** : prévention de l'injection de prompt (séparation contenu / instructions, traitement du contenu externe comme non fiable), moindre privilège pour les agents et les outils, validation humaine des actions à fort impact, journalisation et auditabilité
- Zero trust appliqué à l'IA, IAM et gestion des accès aux modèles et aux données
- Indicateurs de suivi, sensibilisation et montée en compétence des équipes (AI literacy), veille (modèles, vulnérabilités, réglementation).
- Retour d'expérience sur la migration vers une solution d'IA souveraine ou de confiance

Exemples et travaux pratiques :

- Élaboration d'une matrice simple de classification des données pour l'IA
- Identification des mesures de gouvernance et de sécurité à appliquer selon le niveau de sensibilité des données ; revue d'un cas REX (fuite via un service SaaS) et plan de remédiation

Chapitre 6 – Élaborer une stratégie d'IA souveraine et sécurisée

- Grille d'évaluation des solutions d'IA : souveraineté, conformité (RGPD / AI Act / sectoriel), sécurité, coût (TCO incluant GPU et inférence), performance des modèles et réversibilité
- Arbitrages entre exigences de souveraineté, performance des modèles « frontier », contraintes budgétaires, continuité opérationnelle et time-to-market
- Définition des premières étapes d'une trajectoire d'IA souveraine (cas d'usage prioritaires, POC maîtrisé, montée en charge progressive).
- Formalisation d'une feuille de route adaptée aux contraintes de l'organisation

Étude de cas et mise en pratique :

- Étude de cas finale en groupe : simuler une décision stratégique de déploiement d'une IA (souveraine ou de confiance)
    - Analyse du contexte, des contraintes et des risques (juridiques, sécurité, dépendance)

    - Choix argumenté d’une solution ou d’une trajectoire cible
    - Élaboration d’une feuille de route synthétique
    - Restitution collective et discussion sur les conditions de mise en œuvre

Méthodes pédagogiques

- Accompagnement théorique et pratique alliant apports conceptuels, analyses de cas réels (retours d’expérience d’incidents) et mises en situation guidées.
- Le formateur s’appuie sur les expériences et le contexte des participants pour nourrir les échanges de cas concrets et de retours d’expérience ciblés.
- Le format d’une journée privilégie les notions essentielles, les analyses comparatives, les exercices guidés et les cas d’usage directement exploitables.

Les supports de formation seront fournis aux participants au cours de la formation au format PDF.

Méthodes d’évaluation des acquis

Avant la formation :

- Recueil éventuel des attentes et du contexte des participants en amont de la session.

En cours de formation :

- Points d’étapes réguliers du formateur sur la compréhension des stagiaires.
- Questions orales et QCM courts permettant de valider l’acquisition des concepts clés.
- Vérification des résultats lors des travaux pratiques ou exercices guidés.
- Analyse collective de cas d’usage et de retours d’expérience liés à la souveraineté et à la sécurité de l’IA.

Après la formation « à chaud » :

- Questionnaire d’auto-évaluation des compétences complété individuellement par chaque stagiaire en fin de session.
- Questionnaire de satisfaction complété individuellement en fin de session.
- Le compte rendu formateur complété par le formateur.

Après la formation « à froid » :

- Questionnaire de satisfaction complété individuellement quelques semaines après la session.

Un certificat de réalisation de fin de formation est remis au stagiaire lui permettant de faire valoir le suivi de la formation.