

# Formation : IA générative, comprendre, justifier et auditer l'explicabilité des modèles LLM

Expliquez, documentez et auditez les décisions des LLM dans vos projets IA

*Formation pratique - 3j - 21h00 - Réf. LLO*

*Prix : 2360 CHF H.T.*

NEW

Cette formation apporte aux participants les méthodes et outils pour analyser, expliquer et auditer les décisions des LLMs. Vous apprendrez à documenter, justifier et encadrer l'IA générative selon les principes éthiques et réglementaires.

## Objectifs pédagogiques

À l'issue de la formation, le participant sera en mesure de :

- ✓ Comprendre les enjeux stratégiques de l'explicabilité pour l'adoption de l'IA générative.
- ✓ Évaluer les risques métiers liés aux biais, erreurs et opacité des LLM.
- ✓ Concevoir des politiques internes garantissant transparence et responsabilité des usages IA.
- ✓ Déployer des méthodes et outils d'explicabilité pour renforcer confiance et contrôle.
- ✓ Aligner la gouvernance IA avec les cadres réglementaires et les exigences de conformité.
- ✓ Piloter l'audit et la supervision continue des projets GenAI en entreprise.

## Public concerné

Architectes IA, data scientists, PO, juristes, fonctionnels, auditeurs internes et tout professionnel impliqué dans la responsabilité des systèmes IA.

## Prérequis

Connaissances de base sur l'IA ou les LLMs. Pratique de la lecture et de la rédaction de prompts.

### PARTICIPANTS

Architectes IA, data scientists, PO, juristes, fonctionnels, auditeurs internes et tout professionnel impliqué dans la responsabilité des systèmes IA.

### PRÉREQUIS

Connaissances de base sur l'IA ou les LLMs. Pratique de la lecture et de la rédaction de prompts.

### COMPÉTENCES DU FORMATEUR

Les experts qui animent la formation sont des spécialistes des matières abordées. Ils ont été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours qu'ils enseignent. Ils ont au minimum cinq à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

### MODALITÉS D'ÉVALUATION

Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de QCM, mises en situation, travaux pratiques...

Le participant complète également un test de positionnement en amont et en aval pour valider les compétences acquises.

## Modalités d'évaluation

Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de QCM, mises en situation, travaux pratiques...

Le participant complète également un test de positionnement en amont et en aval pour valider les compétences acquises.

## Programme de la formation

### 1 Pourquoi l'explicabilité est essentielle

- Définitions : interprétabilité vs explicabilité.
- Spécificités des modèles génératifs et des LLM.
- Cas d'usage critiques : juridique, médical, RH, finance.
- Risques liés à l'absence d'explication : confiance, adoption, conformité.

### 2 Explicabilité des LLMs, limites et leviers

- Comment fonctionnent les LLM : boîtes noires ou systèmes réductibles ?
- Prompts, mémoire, outputs : où sont les biais ?
- Limitations structurelles : instabilité, hallucination, manque de traçabilité.
- Reproductibilité : challenge ou mirage ?

### 3 Méthodes d'explicabilité dans le contexte GenAI

- Prompt engineering orienté interprétabilité.
- Approches « chain of thought », step-by-step reasoning.
- Justification générée vs preuve calculée.
- Éléments observables dans LangChain : logs, agents, tools.

#### Travaux pratiques

Analyse d'une réponse LLM et justification pas-à-pas. Analyse comparative génération correcte / incorrecte. Reconstruction de la chaîne de raisonnement. Visualisation du contexte et du prompt complet

### 4 Tracer, comprendre et expliquer via LangChain

- Composants traçables dans LangChain : agents, tools, chain logs.
- Logging, callback handlers, prompt templates explicites.
- Introduction à TruLens, PromptLayer, Helicone, LangSmith.
- Création d'un pipeline traçable.

### 5 Utiliser les ontologies et graphes pour expliquer

- Structurer le savoir pour mieux l'expliquer.
- Graphes de connaissance + LLM = contexte interprétable.
- Ontologies métiers : support d'explication intelligible pour l'utilisateur.
- Dialogue entre agent LLM et graphe structuré.

#### Travaux pratiques

Création d'un assistant explicable. Agent qui justifie ses réponses à partir d'un graphe/ontologie. Rédaction de prompts auto-explicatifs. Journalisation complète de la requête à la réponse

#### MOYENS PÉDAGOGIQUES ET TECHNIQUES

- Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : aides audiovisuelles, documentation et support de cours, exercices pratiques d'application et corrigés des exercices pour les formations pratiques, études de cas ou présentation de cas réels pour les séminaires de formation.
- À l'issue de chaque formation ou séminaire, ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques.
- Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le participant a bien assisté à la totalité de la session.

#### MODALITÉS ET DÉLAIS D'ACCÈS

L'inscription doit être finalisée 24 heures avant le début de la formation.

#### ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Pour toute question ou besoin relatif à l'accessibilité, vous pouvez joindre notre équipe PSH par e-mail à l'adresse [psh-accueil@orsys.fr](mailto:psh-accueil@orsys.fr).

## 6 Explicabilité et cadre réglementaire

- Ce que demandent le RGPD et l'IA Act (droit à l'explication, transparence).
- Obligation de documentation, logs, reproductibilité.
- Interfaces explicables : comment afficher une justification intelligible.
- Le rôle de l'explicabilité dans les DPIA et les évaluations de risque.

## 7 Méthodes d'audit des systèmes GenAI

- Créer un log explicatif : prompt + contexte + sources + raisonnement.
- Contrôle qualité des réponses générées (hallucination, cohérence, biais).
- Inclusion d'agents « critique » ou de score d'explication.
- Évaluation humaine des raisonnements.

### Travaux pratiques

Construire un système de réponse auditable. Cas : assistant juridique ou RH.  
Mise en place d'un flux complet avec justification, log d'audit. Démo d'une interface explicative (textuelle + graphique).

## Dates et lieux

### CLASSE À DISTANCE

2026 : 29 juin, 28 sep., 7 déc.