# Formation : Concevoir et industrialiser des architectures RAG avec les LLM

*Formation pratique - 2j - 14h00 - Réf. LLR*

NEW

## Objectifs pédagogiques

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

- Identifier les composants clés d'une architecture RAG et analyser les usages adaptés aux besoins métier.
- Mettre en œuvre un pipeline RAG intégrant ingestion documentaire, vectorisation et génération augmentée.
- Concevoir une architecture RAG sécurisée en tenant compte des contraintes de gouvernance et de conformité. Optimiser et industrialiser un système RAG avec recherche hybride, Reranking et mécanismes d'évaluation.

## Prérequis

Connaissances de base en développement Python ou JavaScript et compréhension générale des API web. Des notions fondamentales en intelligence artificielle, data ou traitement documentaire sont recommandées.

## 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

**PARTICIPANTS**

**PRÉREQUIS**

Connaissances de base en développement Python ou JavaScript et compréhension générale des API web. Des notions fondamentales en intelligence artificielle, data ou traitement documentaire sont recommandées.

**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.

## 1 Comprendre les fondamentaux des architectures RAG

- Identifier les limites des LLM et les apports des architectures RAG dans les usages métier.
- Comprendre les étapes d'un pipeline RAG : ingestion, embeddings, indexation, retrieval et génération.
- Analyser les principes du Chunking, des Embeddings denses et Sparse et de la recherche vectorielle.
- Comparer les approches RAG naïf, avancé et agentique selon les cas d'usage.
- Étudier les tendances récentes : Graph RAG, multi-hop retrieval, Self-RAG et Reflexion.

### Travaux pratiques

Prise en main d'un pipeline RAG simple et analyse du fonctionnement des composants.

## 2 Mettre en œuvre un pipeline RAG opérationnel

- Structurer un pipeline d'ingestion documentaire pour des fichiers PDF et contenus Markdown.
- Réaliser le découpage, la vectorisation et l'indexation de documents dans une base vectorielle.
- Configurer les interactions entre LLM, moteur de recherche vectorielle et génération augmentée.
- Analyser les réponses produites et identifier les mécanismes limitant les hallucinations.
- Utiliser des bases vectorielles telles que FAISS, Chroma, Pinecone ou Weaviate.

### Travaux pratiques

Développement d'un RAG fonctionnel sur un corpus documentaire métier.

## 3 Concevoir une architecture RAG adaptée aux besoins métier

- Formaliser un cas d'usage métier et identifier les attentes des utilisateurs et parties prenantes.
- Évaluer les sources documentaires, la qualité des données et les contraintes de volumétrie.
- Choisir les modèles d'Embeddings, les stratégies de Chunking et les composants d'architecture.
- Concevoir un pipeline d'ingestion cohérent avec les contraintes de performance et d'exploitation.
- Intégrer les enjeux de conformité, de sécurité des données et de gestion des accès.

### Travaux pratiques

Etude de cas et conception d'une architecture RAG complète répondant à un besoin métier.

**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.

## 4 Sécuriser et gouverner un système RAG

- Formaliser un cas d'usage métier et identifier les attentes des utilisateurs et parties prenantes.
- Mettre en œuvre des mécanismes de contrôle des accès et de sécurisation des données sensibles.
- Appliquer les principes de chiffrement des données au repos et en transit dans un environnement RAG.
- Définir des règles de gouvernance pour les usages des LLM et des données internes.
- Évaluer les impacts réglementaires et les exigences de conformité, notamment RGPD.

**Travaux pratiques**
Etude de cas et conception d'une architecture RAG complète répondant à un besoin métier.

## 5 Optimiser les performances et la qualité du Retrieval

- Mettre en œuvre des mécanismes de recherche hybride combinant BM25 et recherche vectorielle.
- Exploiter les techniques de reranking avec cross-encoders pour améliorer la pertinence des réponses.
- Utiliser les filtres de métadonnées et la gestion multi-utilisateurs dans un contexte professionnel.
- Mesurer les performances avec Precision@k, Recall et MRR.
- Identifier les leviers d'optimisation de la qualité et des temps de réponse.

**Travaux pratiques**
Optimisation d'un pipeline RAG avec recherche hybride et évaluation des résultats.

## 6 Industrialiser et déployer un assistant métier basé sur un RAG

- Industrialiser un pipeline d'ingestion avec parsing avancé, nettoyage et enrichissement des données.
- Mettre en place des mécanismes de versioning et de gestion du cycle de vie des données.
- Comprendre les principes des agents et du tool-augmented RAG pour enrichir les interactions.
- Préparer le déploiement et l'exploitation d'un système RAG dans un environnement professionnel.

**Travaux pratiques**
Construction d'un assistant documentaire métier basé sur une architecture RAG avancée.