Construire un assistant documentaire (RAG)

Concevoir et développer en Python un assistant qui répond à partir de vos documents : découpage, embeddings, base vectorielle, génération et évaluation, avec des modèles en API ou hébergés localement.

En pratique

DURÉE
14 h (2 jours)

MODALITÉS
Présentiel · Distanciel · Intra-entreprise · Inter-entreprises

LIEU
Orléans (45), dans vos locaux ou en classe virtuelle

EFFECTIF
3 à 8 participants

TARIF
Inter : 1 290 € HT par participant ; intra : sur devis

DÉLAI D'ACCÈS
30 jours ouvrés entre la demande et le début de la formation, sous réserve d'un nombre suffisant de participants

VALIDATION
Attestation de fin de formation

NIVEAU
Développeur

Objectifs pédagogiques

À l'issue de la formation, le stagiaire sera capable de :

- Expliquer l'architecture d'un système de génération augmentée par la recherche (RAG)
- Préparer et découper un corpus documentaire, calculer des embeddings et les indexer
- Développer une chaîne RAG avec LangChain et un modèle en API ou local
- Évaluer la qualité des réponses et réduire les hallucinations
- Exposer l'assistant dans une interface et en maîtriser les coûts et la confidentialité

Public visé

- Développeurs, data scientists et data engineers
- Équipes techniques chargées de projets d'IA générative

Prérequis

- Pratique de Python (fonctions, bibliothèques, environnements virtuels)
- Notions d'API web (requêtes HTTP, JSON)

Programme

Jour 1 — Les briques d'un RAG

Modèles de langage et API

- Appeler un modèle par API (Mistral, OpenAI, Claude) et en local (Ollama)
- Paramètres, contexte, coûts et limites

Préparer les documents

- Extraction de texte (PDF, Word, pages web)
- Stratégies de découpage et métadonnées

Embeddings et recherche

- Principe des embeddings, choix d'un modèle
- Bases vectorielles (Chroma, pgvector), recherche hybride

Jour 2 — Un assistant fiable et utilisable

La chaîne complète avec LangChain

- Récupération, construction de la requête, génération, citation des sources
- Historique de conversation

Évaluer et améliorer

- Jeux de questions de référence, métriques de pertinence et de fidélité
- Re-classement, filtrage, requêtes reformulées

Mettre à disposition

- Interface avec Streamlit ou Open WebUI
- Sécurité, confidentialité, coûts : bonnes pratiques de déploiement
- Projet final

Méthodes pédagogiques

- Possibilité de travailler sur les données des participants (en intra, après accord)
- Supports de cours, codes source et corrigés remis aux participants
- Formation dispensée par un consultant data et IA qui conçoit et met en production ce qu'il enseigne

Modalités d'évaluation

- Questionnaire de positionnement en début de formation
- Exercices pratiques corrigés tout au long de la formation
- Projet final : assistant documentaire complet sur un corpus fourni, avec grille d'évaluation
- Questionnaire de satisfaction à chaud

Accessibilité

Formation accessible aux personnes en situation de handicap. Contactez-nous en amont (contact@dcid-consulting.fr) afin d'étudier ensemble les adaptations nécessaires.

Financement

DCID Consulting n'est pas certifié Qualiopi à ce jour : cette formation est financée directement par l'employeur et n'est éligible ni au CPF ni à une prise en charge par un OPCO.

Contact et inscription

contact@dcid-consulting.fr · 02 38 25 33 90 ·dcid-consulting.fr/formations/assistant-documentaire-rag/