

# Recommandations de Formation pour Martin (DRH) – Implémentation d'un Mécanisme RAG

## Contexte

Martin, le DRH, est confronté à la difficulté de fournir des recommandations de formation précises pour ses équipes. Les données internes sur les évaluations des employés sont incomplètes ou trop sommaires pour prendre une décision éclairée. Pour résoudre ce problème, vous devez concevoir une solution qui enrichit ces évaluations par la récupération d'informations complémentaires issues d'un corpus documentaire.

## Objectif

Développer une solution qui permet de :

- **Traitement des Évaluations Incomplètes**  
Traiter des données initiales d'évaluation (exemple fourni) qui décrivent de manière sommaire les besoins en formation ou les lacunes de compétences.
- **Récupération d'Informations Complémentaires**  
Implémenter un mécanisme de type **Retrieval-Augmented Generation (RAG)** pour enrichir les données. Cela peut inclure la recherche d'informations dans une base de connaissances ou un corpus documentaire contenant des descriptions de programmes de formation, des meilleures pratiques, des études de cas, etc. Vous êtes libre de construire un index ou d'utiliser des techniques d'indexation (mots-clés, embeddings, etc.).
- **Génération de Recommandations Personnalisées**  
Utiliser un LLM ou une autre méthode de génération de texte pour fusionner les données initiales et les informations récupérées, et ainsi produire un rapport détaillé présentant des recommandations personnalisées pour le développement des compétences.

## Exigences Techniques

### Format d'Entrée Exemple

Même si vous pouvez enrichir le contenu, un exemple de données d'entrée pourrait être sous forme JSON avec des champs basiques, par exemple :

```
{  
  "employe": "Marie Dupont",  
  "evaluation": "Besoins de renforcer la gestion de projet et la  
communication.",  
  "score": 70  
}
```

Vous trouverez la liste complète en Annexe 1.

- **Création d'un Corpus Documentaire**

Vous devez simuler ou intégrer un ensemble de documents (fichiers texte, articles, descriptions de formations, etc.) qui servira de source pour l'enrichissement.

Les données JSON fictives, utilisées pour alimenter votre RAG et répertoriant les sources du corpus documentaire, sont fournies en Annexe 2

- **Architecture Multi-Nœuds**

Votre chaîne doit comporter plusieurs étapes, dont au moins :

1. **Ingestion et Prétraitement des Évaluations**
2. **Récupération d'Informations depuis le Corpus**
3. **Génération de Recommandations**

La manière dont vous enchaînez ces étapes est à définir librement.

- **Tests Unitaires**

Ajoutez des tests unitaires pour vérifier le fonctionnement de chaque nœud de votre chaîne.

- **Documentation**

Commentez le code et/ou fournissez un README expliquant vos choix d'implémentation, notamment sur la mise en œuvre du mécanisme RAG.