

Projets Retrieval-Augmented Generation (RAG)

Idées pour le projet :

1. Assistant de Recherche Documentaire

- **Difficulté :** ★★★★★
- **Améliorations :**
 - Ajouter un système d'indexation des documents (vectorisation avec FAISS, Elasticsearch, ou Weaviate).
 - Comparer différentes méthodes de récupération d'information (BM25 vs embeddings).
 - Tester plusieurs modèles RAG et évaluer leurs performances (ex. LlamaIndex vs Haystack).

2. Chatbot pour FAQ d'Entreprise

- **Difficulté :** ★★★★★
- **Améliorations :**
 - Ajouter un module d'apprentissage automatique pour affiner le modèle sur des données internes.
 - Permettre l'ajout dynamique de nouvelles FAQ et tester un système de feedback utilisateur.

3. Moteur de Recherche Médicale

- **Difficulté :** ★★★★★
- **Améliorations :**
 - Intégrer un pipeline NLP pour l'extraction d'entités nommées (NER) sur les termes médicaux.
 - Vérifier la fiabilité des sources et inclure des indicateurs de confiance.

4. Assistant de Programmation Augmentée

- **Difficulté :** ★★★★★
- **Améliorations :**
 - Ajouter un composant de complétion de code basé sur les résultats trouvés.
 - Utiliser OpenAI Codex ou Code Llama en complément pour générer du code contextuel.

5. Système de Recommandation pour Bibliothèque Numérique

- **Difficulté :** ★★★★★

- **Améliorations :**
 - Intégrer un score de pertinence basé sur des embeddings et des similarités cosines.
 - Expérimenter différentes stratégies de résumé (ex. extractif vs abstractive).
-

Autres idées de projets RAG :

Générateur de résumés intelligents pour articles de recherche

- **Objectif :** Extraire et résumer automatiquement des articles en mettant en avant les points clés.
- **Difficulté :** ★★★★★
- **Technos :** LangChain, LlamaIndex, OpenAI, FAISS

Assistant Juridique Augmenté

- **Objectif :** Répondre à des questions juridiques en se basant sur des bases de données légales.
- **Difficulté :** ★★★★★
- **Technos :** Elasticsearch, GPT-4, Corpus OpenLaw

Outil d'Analyse de Sentiments sur Documents Financiers

- **Objectif :** Extraire des tendances et des insights depuis des rapports annuels ou des articles économiques.
 - **Difficulté :** ★★★★★
 - **Technos :** NLP + RAG, API Edgar SEC, Pinecone
-

Organisation des projets RAG

- **Travail en groupe de 2-4 étudiants**
- **Livrables attendus :**
 1. **Code fonctionnel** du projet, incluant le modèle RAG et l'intégration avec une base documentaire.
 2. **PowerPoint de 6 slides** comprenant :
 - Slide 1 : Introduction et objectif du projet
 - Slide 2 : Présentation du concept de RAG appliqué au projet
 - Slide 3 : Architecture technique (schéma du pipeline de récupération et génération)
 - Slide 4 : Technologies utilisées (LLM, base vectorielle, etc.)

- Slide 5 : Démonstration du fonctionnement (exemples de requêtes/réponses)
 - Slide 6 : Conclusion et perspectives d'amélioration
3. **Démonstration interactive avec Streamlit**, illustrant l'idée de mise en œuvre du RAG.
- Interface permettant aux utilisateurs de poser des questions et voir les réponses générées.
 - Visualisation des documents récupérés en backend.

Évaluation :

- Qualité de l'implémentation et pertinence des résultats.
- Clarté de la présentation et démonstration.
- Justification des choix techniques.