

PLMIND : Développement d'un chatbot interne basé sur un LLM et une architecture RAG

Contexte

Le site intranet d'un laboratoire contient une grande quantité d'informations essentielles au fonctionnement scientifique et administratif : procédures, consignes, formulaires, documents PDF, etc.

Cependant, ces contenus sont dispersés et leur recherche est souvent laborieuse.

Dans le cadre d'un projet prospectif, ce projet a pour but d'explorer la mise en place d'un **chatbot local** reposant sur un **modèle de langage (LLM)** et une **architecture RAG (Retrieval-Augmented Generation)**, afin de faciliter l'accès à ces informations sans compromettre la confidentialité des données internes.

Objectif du projet

Concevoir un **prototype fonctionnel** de chatbot capable d'interroger les contenus du site interne (HTML et PDF) du LIP6 et de fournir des réponses **contextualisées et sourcées** (avec lien vers la source d'origine).

Le projet devra également inclure une **analyse et justification des choix techniques et scientifiques** réalisés, dans une optique exploratoire.

Travaux attendus

Le projet sera structuré en plusieurs étapes :

1. **Analyse du besoin et exploration des solutions existantes** (LLM open source, bibliothèques RAG, outils d'indexation).
2. **Mise en place d'un pipeline d'ingestion et d'indexation** des documents internes (HTML, PDF).
3. **Développement d'un prototype de chatbot** (backend + interface minimale d'interaction).
4. **Évaluation des performances et de la pertinence des réponses.**
5. **Rédaction d'un rapport final** incluant une **discussion sur les choix techniques et méthodologiques** et une **notice d'installation** du système.

Encadrement

- **Encadrant(s) universitaire(s)** : *Jean-Noël VITTAUT (LIP6/LFI) & Pierre-Henri WUILLEMIN (LIP6/Decision)*

Compétences mobilisées

- Programmation en **Python**.
- Connaissances en **traitement automatique du langage (NLP)** et en **LLM**.

- Bases en **gestion documentaire** (HTML, PDF, indexation).
- Capacité à **structurer une démarche scientifique** et à **analyser des choix techniques**.

Organisation et livrables

- Projet logiciel MIND réalisé par **2 étudiants M1 du Master MIND**.
- Livrables attendus :
 - Prototype fonctionnel du chatbot.
 - Pipeline d'indexation et base de connaissances locale.
 - Documentation technique et notice d'installation.
 - Rapport final de synthèse (méthodes, choix techniques, perspectives).
 - Présentation orale en fin de projet.

Apports pédagogiques

Ce projet permettra aux étudiants de :

- Mettre en pratique des compétences en **intelligence artificielle, NLP et développement logiciel**.
- Découvrir les enjeux liés à la **confidentialité et à l'hébergement local des LLM**.
- Participer à un **projet prospectif de recherche appliquée** au sein d'un laboratoire scientifique.