

Projet NLP (LLM)

Note :

- Les projets doivent être soumis par mail au plus tard le 20 avril 2025, sous forme d'un fichier compressé à l'adresse a.sadiq@uit.ac.ma.
- Le fichier devra contenir un rapport détaillant les étapes de réalisation du projet, ainsi que le code au format Jupyter Notebook avec une explication de chaque étape.
- Chaque projet doit être réalisé en binôme ou monôme.

1. Assistant de Synthèse d'Articles de Recherche

L'objectif est de développer un outil permettant aux étudiants de résumer des articles scientifiques en quelques phrases.

Technologies à utiliser : GPT-4, LangChain, Pinecone, Flask, React

Source des données :

- ArXiv API (<https://arxiv.org/help/api/>)
- HAL archives ouvertes (<https://hal.archives-ouvertes.fr/>)

Les tâches :

- Back-end & LLM:
 - Implémentation de l'extraction des articles via l'API ArXiv et HAL
 - **Intégration de LangChain pour interagir avec GPT-4**
 - **Création du pipeline de résumé automatique avec mise en avant des concepts clés**
 - Stockage des résumés dans Pinecone pour un accès rapide
- Front-end:
 - Développement de l'interface en React avec un champ de recherche
 - **Connexion à l'API Flask pour récupérer les résumés générés**

2. Générateur Automatique de QCM à partir de Cours

L'objectif est de transformer des supports de cours en questionnaires à choix multiples.

Technologies à utiliser : GPT-4, LangChain, ChromaDB, Flask

Source des données :

- Cours en PDF/Word (cours de votre choix)
- OpenCourseWare (<https://ocw.mit.edu/>)

Les tâches :

- Back-end & LLM:
 - Implémentation du chargement et du traitement des fichiers PDF/Word
 - Extraction des concepts clés avec LangChain et GPT-4
 - Génération automatique des questions et des réponses
 - Stockage des QCM générés dans ChromaDB
- Front-end:
 - Développement de l'interface Flask pour uploader des documents
 - Création d'un tableau d'affichage des questions générées
 - Ajout d'une option pour exporter les QCM en PDF