

# Rapport détaillé du projet d'Application Flask avec IA

## 1. Objectif du projet

Développer une application web (Flask) permettant de : - Charger un fichier PDF - Générer un résumé intelligent du contenu - Générer des questions ainsi que des réponses à la demande

## 2. Architecture du projet

- **app.py** : Point d'entrée de l'application (routes HTTP)
- **backend/pipelines** : Contient le cœur du traitement du texte
- **pdf\_pipeline.py** : Extraction du texte brut du PDF
- **generation\_pipeline.py** : Génération du résumé via Transformer
- **rag\_pipeline.py** : Recherche de réponses pertinentes
- **save\_json.py** : Export des résultats
- **backend/model** : Fichiers du modèle (vocab.json, merges.txt)
- **templates** : Pages HTML (index, result, quiz, error)
- **uploads** : Fichiers déposés par l'utilisateur

## 3. Technologies utilisées

- Framework : Flask
- Extraction de texte : PyMuPDF (fitz)
- Tokenization : ByteLevelBPETokenizer
- Génération de texte : PyTorch + Transformer
- Recherche de réponses : SentenceTransformers
- Détection de langue : Langdetect
- Décodage du texte généré : Nucleus Sampling

## 4. Problèmes rencontrés

- Chemins relatifs du modèle (solution : chemins absolus via os.path.join)
- Langue du texte (solution : traduction si texte en français)
- Qualité du résumé généré (cause : modèle peu adapté, généraliste)
- Limites mémoire du modèle (solution : forçage du calcul en mode CPU)

## 5. Ce qui fonctionne

- Upload de fichier
- Extraction du texte
- Génération du résumé
- Génération de questions spécifiques (RAG)

## 6. Améliorations suggérées

- Ajouter une barre de progression côté frontend

- Limiter la taille des PDFs acceptés
- Améliorer la précision du modèle de résumé (modèle extractif ou français)
- Optimiser l'usage mémoire (modèle léger/hébergé)
- Ajouter des métadonnées spécifiques du PDF

## **7. Conclusion générale**

L'application repose sur une architecture complète intégrant des modèles NLP ainsi qu'une couche applicative robuste. Elle répond techniquement à la majeure partie des besoins définis, tout en nécessitant des améliorations significatives pour atteindre des résultats plus précis ainsi qu'une utilisation optimale des ressources mémoire.

Si nécessaire, je peux proposer une version Markdown du document ainsi qu'une liste de modèles de résumé adaptés. Souhaitez-vous en disposer ?