**Frequently Asked Questions**

**Q1 : Comment avez-vous collecté les données de départ pour votre projet ?**

**R1 :** Au début du projet, on a tenté de collecter nos propres données en utilisant les métadonnées d'ArXiv. On a réussi à rassembler une base de données de fichiers PDF à partir de ces métadonnées. Cependant, le contenu de ces PDF était long et difficile à mnipuler, ce qui a posé des problèmes de traitement.

**Q2 : Comment avez-vous géré le problème de traitement des PDFs complexes ?**

**R2 :** Pour résoudre le problème de traitement des PDFs complexes, on a opté pour une approche différente. On a utilisé la base de données SQuAD (Stanford Question Answering Dataset) pour l'entraînement du modèle BERT de Hugging Face spécialement conçu pour l'extraction de réponses à partir de textes complexes.

**Q3 : Avez-vous rencontré des défis particuliers lors de la mise en place de votre chatbot ?**

**R3 :** Oui, l'un des défis était de gérer les cas où les modèles ne pouvaient pas extraire de réponses précises en raison de la complexité du texte. Pour résoudre cela, on a dû travailler sur l'amélioration de la qualité des réponses et sur la manière de gérer les cas où aucune réponse adéquate n'était trouvée.

**Q4 : Qu'est-ce que cette application fait exactement ?**

**R4 :** Cette application analyse des PDF d'articles scientifiques et extrait des informations pertinentes à l'aide d'un chatbot. Elle simplifie la recherche et la récupération de données importantes à partir d'articles académiques.

**Q5 : Quelles informations peut-on extraire des articles scientifiques ?**

**R5 :** L'application peut extraire diverses informations, y compris les titres, les auteurs, les résumés, les citations, les références, les données expérimentales, les résultats, etc.

**Q6 : De quelles bases de données ou sources d'articles scientifiques cette application peut-elle extraire des données ?**

**R6 :** Notre application est capable d'analyser des articles provenant de diverses sources et bases de données, y compris PubMed, IEEE Xplore, ScienceDirect, etc. Elle est également capable de traiter des PDF personnalisés si vous les téléchargez.

**Q7 : Quelles ont été les principales leçons que vous avez apprises lors de ce projet ?**

**R7 :** Ce projet nous a appris l'importance de choisir les bonnes sources de données pour l'entraînement de modèles NLP. On a également appris à gérer les défis liés à la complexité du texte scientifique. En outre, on a amélioré nos compétences en ingénierie logicielle en développant une interface utilisateur conviviale avec Streamlit.

**Q8 : Quels sont les domaines de l'intelligence artificielle et du traitement du langage naturel (NLP) que vous avez explorés dans ce projet ?**

**R8 :** Notre projet a exploré les domaines de la classification de texte, de l'extraction d'informations et de la génération de réponses automatiques, tous relevant de l'intelligence artificielle et du NLP.

**Q9 : Avez-vous des recommandations pour d'autres personnes travaillant sur des projets similaires dans le domaine de l'analyse d'articles scientifiques et de l'extraction de données ?**

**R9 :** Oui, on recommande de soigneusement sélectionner les bases de données et les modèles NLP appropriés en fonction des besoins spécifiques du projet. Il est également important de bien documenter le processus d'entraînement des modèles et de continuer à suivre les évolutions dans le domaine de l'IA et de l'apprentissage automatique.

**Q10 : Quelles étaient vos attentes initiales avant de commencer le stage, et ont-elles été satisfaites ?**

**R10 :** Initialement, on s'attendais à rencontrer des défis techniques dans le traitement de données complexes, et ces attentes se sont réalisées. Cependant, grâce à des approches innovantes, on a pu surmonter ces défis avec succès.