**Rapport de Tests et Évaluations des Performances**

1. Introduction

L'objectif de cette section est d'évaluer les performances des réponses générées par le modèle Gemini-1.5-Flash et de les comparer avec des réponses obtenues via TF-IDF. Deux métriques principales sont utilisées : la complétude et la fiabilité

2. Métriques Utilisées

2.1 Complétude

* **Définition** : La complétude est évaluée en demandant au modèle de donner un score entre 0 (incomplet) et 1 (complet).  
  Une réponse complète couvre tous les éléments essentiels mentionnés dans la réponse de référence.

2.2 Fidélité

* **Définition** : La fidélité est mesurée en comparant le texte généré avec une source de vérité (ex : documents récupérés).  
  Un score de 1 signifie que la réponse est totalement fidèle à la source, tandis qu’un score proche de 0 indique des divergences.

Cas d'étude :  
Question testée : "Comment avoir mon passeport ?"  
Constat : La réponse générée par le modèle mentionne certains documents nécessaires à la demande, mais n'est ni complète ni totalement fidèle aux sources officielles.

Résultats pour la question "Comment avoir mon passeport ?"

* Complétude (Score : 0.60)
  + Explication : Le score de 0.60 indique que la réponse est partiellement complète.
  + Analyse : La réponse mentionne des documents importants, mais manque de détails cruciaux (exemple : durée du processus, lieux exacts, types de passeports, tarifs).
  + Interprétation : La réponse couvre partiellement les informations requises ; des informations supplémentaires et plus précises sont nécessaires.
* Fidélité (Score : 0.20)
  + Explication : Le score de 0.20 reflète une faible fidélité à la source.
  + Analyse : La réponse, bien que logique, manque de précision et comporte des informations qui ne proviennent pas forcément de sources officielles (exemple : tarif de 30 000 CFA susceptible de varier).
  + Interprétation : Le modèle n'a pas fourni une réponse pleinement fidèle aux documents ou sources officielles, nécessitant ainsi des améliorations pour garantir la précision.

3. Comparaison avec d'Autres Modèles ou Approches

Les tests ont été réalisés sur les trois questions suivantes :

1. Comment faire son CIP ?
2. Comment obtenir un passeport ?
3. Quels sont les documents nécessaires pour une demande de passeport ?

Tableau récapitulatif des scores

| Question | Complétude (Gemini) | Fidélité (Gemini) | Complétude (TF-IDF) | Fidélité (TF-IDF) |
| --- | --- | --- | --- | --- |
| Comment faire son CIP ? | 0.7 | 0.4 | 0.2 | 1.0 |
| Comment obtenir un passeport ? | 0.6 | 0.6 | 0.2 | 0.2 |
| Quels sont les documents nécessaires pour un passeport ? | 0.6 | 0.4 | 0.5 | 0.3 |

4. Explication des Résultats

4.1 Analyse par Question

Question 1 : Comment faire son CIP ?

* Gemini
  + Complétude : 0.7  
    La réponse est relativement complète, couvrant les principaux aspects du processus pour faire un CIP, même si certains détails peuvent manquer.
  + Fidélité : 0.4  
    La réponse est partiellement fidèle, avec possiblement des étapes ou précisions manquantes.
* TF-IDF
  + Complétude : 0.2  
    La réponse est très partielle, ne couvrant pas l'ensemble des facettes de la question.
  + Fidélité : 1.0  
    La réponse correspond exactement à l’intention de la question, même si elle manque de détails.

Question 2 : Comment obtenir un passeport ?

* Gemini
  + Complétude : 0.6  
    La réponse est assez complète, bien que certaines informations supplémentaires puissent être nécessaires.
  + Fidélité : 0.6  
    La réponse correspond globalement à la question, avec une fidélité modérée.
* TF-IDF
  + Complétude : 0.2  
    La réponse est encore plus partielle que celle de Gemini.
  + Fidélité : 0.2  
    Le score faible montre que la réponse n’adhère pas entièrement à l’intention de la question.

Question 3 : Quels sont les documents nécessaires pour une demande de passeport ?

* Gemini
  + Complétude : 0.6  
    La réponse est modérément complète, couvrant les principaux documents nécessaires, mais des détails peuvent manquer.
  + Fidélité : 0.4  
    La réponse est partiellement fidèle et pourrait être affinée selon les contextes ou exceptions.
* TF-IDF
  + Complétude : 0.5  
    La réponse est plus complète que pour la question précédente, bien qu’elle reste partielle comparée à Gemini.
  + Fidélité : 0.3  
    Le score relativement bas indique que la réponse n’est pas entièrement fidèle à la question.

5. Interprétation des Résultats

5.1 Complétude

* Gemini :  
  Excelle en complétude avec des scores nettement plus élevés, couvrant une plus grande partie des informations demandées, offrant ainsi des réponses plus détaillées.
* TF-IDF :  
  Privilégie la fidélité au détriment de la complétude, fournissant des réponses très précises mais souvent incomplètes.

5.2 Fidélité

* TF-IDF :  
  Se distingue par sa capacité à fournir des réponses parfaitement fidèles à la question posée, notamment pour des questions précises comme "Comment faire son CIP ?". Cependant, ce niveau de fidélité s'accompagne d'un manque de détails.
* Gemini :  
  Bien que ses réponses soient plus complètes, la fidélité est moins élevée, suggérant un modèle plus général qui peut manquer de précision dans certaines situations.

6. Conclusion et Recommandations

* Gemini :
  + Plus adapté pour des questions nécessitant des réponses complètes.
  + Peut manquer de précision sur certains points.
* TF-IDF :
  + Plus performant pour garantir la fidélité et la précision.
  + Tendance à négliger certains aspects de la question, menant à une réponse incomplète.

Recommandation :  
En fonction des objectifs du projet, il peut être pertinent d'utiliser l'un ou l'autre de ces modèles, voire une combinaison des deux afin de maximiser à la fois la complétude et la fidélité des réponses.

* Pour des cas où la précision et la concision sont essentielles, TF-IDF est préférable.
* Pour des cas nécessitant des réponses détaillées et complètes, Gemini est plus adapté.