

)



## Conclusion

### Cas d'utilisation du projet :

- Lecture automatique et repérage de CV par mots clés permettant un tri rapide et efficace des candidatures (étape coûteuse en termes de temps dans les processus de recrutement)
- Lecture automatique de note de frais: Détection de ligne comptables
- Détection de montant - Détection de TVA
- Traitement documentaire dans le cadre de la gestion éditique de document. Cela permet l'enregistrement de data en base de données afin d'améliorer la traçabilité et la recherche

Traçabilité/décryptage de documents d'époques : aide pour décrypter les documents anciens/abimés par le temps.

### Perspectives d'amélioration du code :

- Nous avons donc un modèle qui répond à notre problématique. Celui-ci détecte le texte en alphabet latin et le reconnais.
- Le modèle n'inclut pas de caractères spéciaux et la ponctuation. Pour inclure ces caractères, une possibilité serait d'augmenter le dataset en incluant plus de caractères spéciaux. L'autre possibilité serait de créer un modèle spécifique prenant en compte ces caractères et que ce modèle soit additionné au modèle développé dans notre projet ...
- Pour améliorer les performances, on pourrait aussi augmenter les données (91 992 images ne suffisent pas forcément), ou réaliser la même méthode mais en partant sur un apprentissage des phrases/expression plutôt que les mots, voir additionner les 2 méthodes.
- Une des grandes difficultés de ce type de problématique vient directement du poids des données utilisées et de la puissance de calcul disponible pour entraîner nos modèles sur des machines non

- dimensionnées pour ça. Une des pistes d'améliorations possible de notre projet en travaillant sur un meilleur traitement des images en termes de performance et de passer par des services comme google collab pour aller chercher plus de puissance de calcul.
- Enfin une dernière piste serait de se servir des règles de la langue anglaise. On pourrait recenser les mots les plus utilisés dans la langue et ainsi associé à la reconnaissance une probabilité d'avoir tel mot. On pourrait à partir de correcteur de grammaire, et d'orthographe, rajouter une probabilité que le mot à reconnaître soit d'une telle nature et ainsi filtrer les choix de reconnaissance. Cependant toutes ces pistes nécessitent énormément de ressources de calcul et pourrait ne pas être viable.

Made with Streamlit