**Projet fédérateur IA**

Sujets

1. **(Aybuke) construire le site technologique de l’IPSA (agrégateur d’infos)**
2. (Aybuke) Drone-Based attendance tracking tool
3. (Othmane) Web application for furniture recognition and price tagging via AI and Google Lens API
4. (Othmane) Job/Internship Scraping Platform with AI-powered chatbot

Reprendre l’algo en data mining en y automatisant la collecte des sites et en permettant toutefois à l’humain de vérifier la véracité des informations collectées.

Recherches BERT :

BERT (Bidirectional Encoder Représentations from Transformers) : modèle de langage développé par Google en 2018 pour le NLP

Basé sur le Transformeur, n’utilisant que sa partie « encodeur »

Modèle pré-entraîné pour 2 objectifs :

* Masked Language Modeling : une des unités de la suite est remplacée par l'unité [MASK]. L'objectif est alors que la distribution de probabilité en sortie du modèle maximise la probabilité de prédiction de l'unité masquée.
* Next Sentence Prediction : la suite en entrée est composée de deux phrases. Le modèle doit prédire (*vrai* ou *faux*) si les deux phrases sont successives dans les données d'entrainement ou non.

2 versions pré-entraînées : base (12 blocs encodeurs, 12 têtes d’attention et 110 millions de paramètres) et large (24 blocs, 16 têtes et 340 millions de paramètres)

Possibilité d’implémentation en Python avec Tensorflow