Stage Développement IA

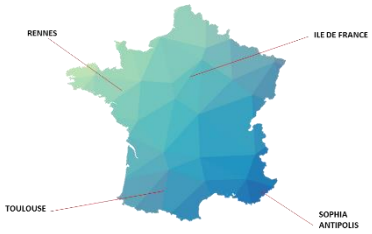
Exploitation de l'IA pour une solution de détection de données personnelles

ALTEN Rennes – Stage de fin d'étude (PFE) 6 mois – Bac+5

Qui sommes-nous ?

ALTEN est une entreprise de conseil en ingénierie. Fort de nos 34 000 collaborateurs et de notre implantation dans plus de 28 pays, nous apportons des solutions innovantes pour contribuer aux projets et au développement de nos clients issus de secteurs d'activité variés. Faire d'ALTEN "la Maison de l'ingénieur", c'est l'ambition de notre Groupe qui compte parmi les ingénieurs de talent qui le rejoignent chaque année, plus de la moitié de jeunes diplômés. Labélisé en 2019 « Top employeur » pour la 8ème année consécutive, nous sommes un acteur majeur du recrutement d'ingénieurs en France et dans le monde.

Faire un stage dans nos Lab



Cellules souples et agiles, ALTEN a mis en place plusieurs « Labs » au sein de la Direction de l'Innovation, avec l'objectif de favoriser les projets innovants en réponse aux enjeux de ses clients, et leur fournissant des solutions industrielles originales voire disruptives dans l'accompagnement de leur R&D.
Le Lab Ouest situé sur Rennes accueille dans ses équipes des stages de fin d'étude sur les domaines suivants : la **cybersécurité**, la **blockchain**, l'**IA**, le **Machine Learning**, le traitement en langage naturel (**NLP**), l'**automatisaiton** des tests.

Présentation du stage

Contexte

L'anonymisation est le fait de purger les données d'une entité de toutes données sensibles ; c'est à dire des données directement ou indirectement liées à une personne physique identifiée et identifiable.

L'anonymisation répond entre autre à un besoin de mise en conformité pour la RGPD.

Objectifs ALTEN

A ce titre ALTEN a décidé de développer une boite à outils qui réponde aux différents aspects de l'anonymisation. Le premier besoin est de pouvoir à partir d'un document détecter si oui ou non, il contient des données personnelles.

Après un état de l'art sur les approches et les solutions possibles, le but est de concevoir une solution qui à partir d'un document type pdf, soit capable de dire avec quel niveau de certitude le document contient des données personnelles.

MOTS·CLÉS

IA - NLP

Réseau de neurones, embedding, NER, Transformers…
Pytorch, Tensorflow
POS Tagging
Python

Contact : maud.amice@alten.com