Stage Intelligence Artificielle – NLP

Objectif: **ORIENTATION DES USAGERS VERS LES SERVICES DE PLAINTES EN LIGNE**

Domaine: Intelligence Artificielle (**IA**) / Natural Language Processing (**NLP**) / Natural langage Understanding (**NLU**)

Structure d'accueil

L'ANFSI est chargée du développement, de la mise en œuvre, de la sécurité des systèmes d'information, des équipements numériques et des applications au profit des forces de sécurité intérieure (gendarmerie / police nationale). Le datalab de l'ANFSI est une structure de 15 personnes dédiée à la valorisation des données internes à travers l'intelligence artificielle. Le stagiaire sera accueilli au sein du datalab.

Contexte métier & objectif

Il existe plusieurs services du ministère de l'intérieur pour recueillir les plaintes de citoyens. Chaque service est dédié à un périmètre d'infractions précis. Lorsqu'une victime souhaite déposer plainte, il est probable qu'elle passe par un moteur de recherche pour trouver comment faire. Selon la requête et le moteur employé, elle peut arriver sur différents sites du ministère.

A partir d'une première description succincte des faits par la victime, et éventuellement certains champs structurés, nous souhaitons l'orienter automatiquement vers un des 6 canaux actuellement disponibles pour recueillir des plaintes (*Plainte en ligne, Thésée, Perceval,...*)

Éléments techniques

On pourra envisager un chatbot succinct (i.e nombre de tours de paroles limités), ou un classifieur immédiat qui fournit une réponse directement à partir des inputs. Le choix de la solution se fera à partir de l'analyse des données historiques afin d'étudier les informations spontanément fournies par les victimes, et si elles permettent de prendre une décision.

Le classifieur pourra être conçu suivant différents paradigmes :

- **Option 1 : extraction d'information** suivie de règles métier
- **Option 2 :** modèle de **classification de textes** end-to-end

- **fine-tuning de LLM**
- **injection de contexte (prompt engineering)**

Compte-tenu de la nature du problème, une réflexion devra être menée sur **l'explicabilité** de l'algorithme implémenté et les éventuels **biais** associés.

Fonctions / Tâches:

Au sein du Datalab, le stagiaire se verra confier les missions suivantes:
- Analyse des données historiques
- Réalisation d'un état de l'art pour évaluer quelles sont les pistes d'améliorations les plus prometteuses. Eventuellement suggestions de nouvelles pistes que celles évoquées ci-dessus.
- Implémentation d'un modèle
- Mesure des performances
- Analyse critique des erreurs

Profil du candidat:

REQUIS

- Étudiant en dernière année d'école ingénieur ou Master 2, dans un parcours IA / data science.
- Première expérience en NLP dans le cadre des études
- Première expérience en Deep Learning dans le cadre des études
- Maîtrise du français (courant) et de l'anglais (technique)

SOUHAITE

- Connaissance du langage Python
- Connaissance des outils : pytorch, langchain, Dataïku
- Expérience avec des modèles de type TRANSFORMER

Durée et dates indicatives du stage	Stage de 6 mois, date de début à discuter
Organisme d'accueil	Direction générale de la gendarmerie nationale Agence du numérique des forces de sécurité intérieure Datalab
Localisation	4 rue Claude Bernard 92130 ISSY LES MOULINEAUX
Tuteur de stage	Nom et prénom: PERTSEKOS Daphné Fonctions: Data Scientist
Contact	daphne.pertsekos@gendarmerie.interieur.gouv.fr
Conditions	- durée hebdomadaire de présence : 35 heures – temps complet - accès au restaurant administratif - accès à la salle de sport - prise en charge 50 % des dépenses Navigo (transport lieu d'hébergement / lieu de stage) - gratification 4,35€/H