

Compte rendu personnel de :  
**LANNUZEL Dylan**

Manager actuel : **Caillaud Maxime**

## **Projet Finance GPT - Compte rendu n°1**

**du 06/11/2023 au 17/11/2023**

<b>Tâches effectuées</b>	<b>Durée</b>
Serveur Discord	1h
Recherche d'informations Analyse juridique	2h
Prise en main GPT-2 (Recherche et implémentation)	8h

### **Objectif sur cette période :**

- Se renseigner sur l'aspect juridique du dataset
- Comprendre et implémenter le modèle pré-entraîné GPT2

### **Tâches accomplies**

- Serveur Discord :

J'ai mis en place en coopération avec mon équipe un serveur discord clair et organisé avec l'utilisation de divers canaux et post type forum. Cela nous permet d'être efficace dans l'organisation de nos ressources et de notre répartition de travail.

- Recherche d'informations Analyse juridique :

Avec l'utilisation de données, il a fallu nous renseigner sur les conditions d'utilisation du jeu de données. Pour ce faire, j'ai pu étudier les caractéristiques juridiques de la licence associée. Ce travail a permis ensuite de rédiger l'analyse juridique.

- Prise en main GPT-2

J'ai pu étudier les diverses approches permettant de pouvoir importer et utiliser le modèle

pré-entraîné GPT-2. A la suite de cela, j'ai pu lui passer en entrée du texte et voir la 'magie' opérer.

J'ai vraiment trouvé cette partie du travail intéressante et m'a conclu dans mon idée que l'IA est vraiment à la portée de tous.

## Difficultés rencontrées :

- Multiples duplication du modèle :
  - Aux premières itérations du modèle, j'ai constaté avec Maxime beaucoup de duplication notamment à la fin où le modèle essayait de combler la taille maximale donnée en répétant les derniers mots.
  - Pour résoudre ceci, j'ai pu ajuster le paramètre `no_repeat_ngram_size=3` et `temperature=0.8` pour pouvoir fixer le choix des séquences à minimum 80% de pertinence et éviter de répéter plusieurs fois trois mots à la suite.

## Conclusion :

Ce début de travail a été pour moi un plongeon dans un contexte IA dont je n'ai pas eu l'habitude jusqu'à aujourd'hui et m'a permis de rapidement cerner les enjeux du modèle GPT-2 et de ses spécificités. Je suis plus que jamais motivé à mener à bien ce projet.