

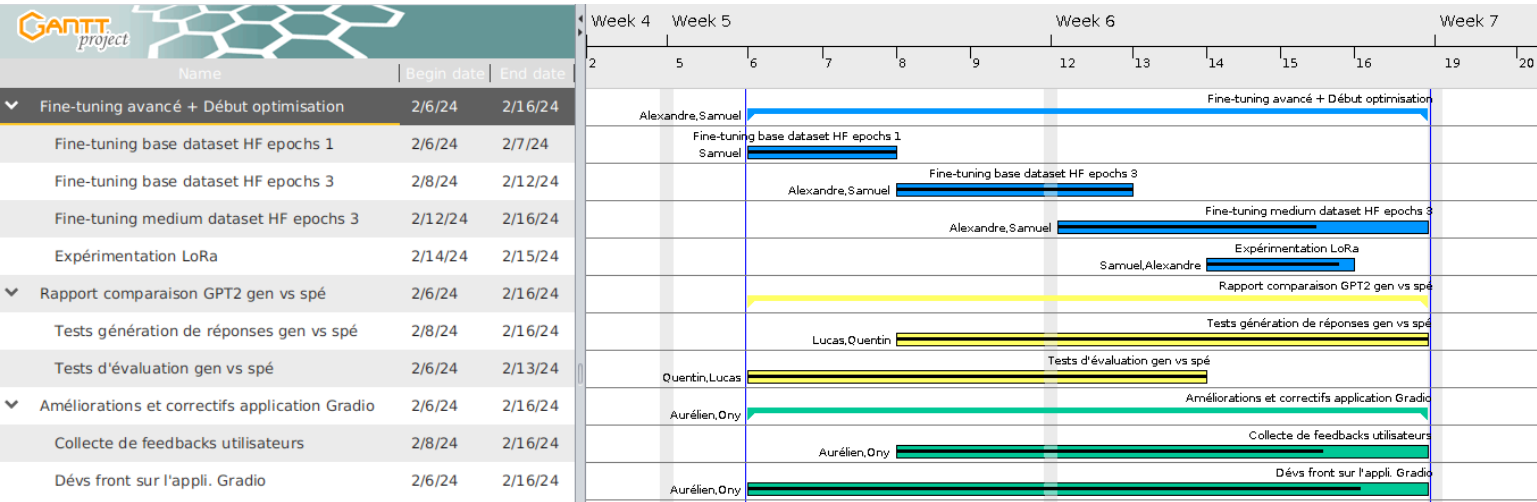
# SAÉ 5.A.01 - Rapport de management

## Période 05/02/2024 - 16/02/2024

Durant les semaines du 05/02 au 16/02/2024, j'ai été manager pour le livrable n°7 consistant à la réalisation d'un rapport de comparaison entre le modèle GPT-2 générique et spécialisé. Notre groupe a traité plusieurs sujets en parallèle, en débutant notamment l'optimisation du modèle via la réalisation de nouveaux fine-tunings, mais également l'amélioration et les corrections du prototype de l'application et de son Interface Homme-Machine.

L'équipe a été réparti en trois groupes pour traiter ces diverses tâches :

Groupe 1 <u>Fine-tuning avancé et début d'optimisation</u>	Groupe 2 <u>Rapport de comparaison entre le modèle générique et spécialisé</u>	Groupe 3 <u>Améliorations et correctifs sur le prototype de l'application Gradio</u>
<div>- Samuel DORISMOND</div> <div>- Alexandre HUYNH</div>	<div>- Lucas AGUETAÏ</div> <div>- Quentin VERMEERSCH</div> <div>- Alexandre H. (en support)</div>	<div>- Aurélien ZUFIC</div> <div>- Ony ANDRIATSAHAVOJAONA</div> <div>- Samuel D. (en support)</div>



(Gantt disponible en annexe à ce rapport)

### Groupe 1 : Fine-tuning avancé et début d'optimisation

- Fine-tunings modèles GPT-2 base / medium / large, paramètres 1 à 3 epochs
  - Ré-entraînement en utilisant la dataset "scientific-papers" de Hugging Face
    - identique à celle de Tensor Flow téléchargée et divisée manuellement (méthode utilisée employée précédemment), mais plus pratique et moins important en consommation mémoire
  - Entraînement du modèle sur GPT-2 base et medium en epochs 1 et 3
- Etude et expérimentation de la technique LoRa

### **Groupe 2 : Rapport de comparaison entre le modèle générique et spécialisé**

- Tests génération texte modèle générique vs spécialisé
  - Modèles spécialisés utilisés pour la comparaison
    - GPT-2 base (dataset sci.-papers TF avec split manuelle en fichiers)
    - GPT-2 base (dataset sci.-papers HF, epochs 1)
    - GPT-2 base (dataset sci.-papers HF, epochs 3)
- Tests d'évaluations performance : précision et taux d'échec / pertes
  - Modèles spécialisés utilisés pour la comparaison
    - GPT-2 base (dataset sci.-papers HF, epochs 1)
    - GPT-2 base (dataset sci.-papers HF, epochs 3)

### **Groupe 3 : Améliorations et corrects front-end sur le prototype de l'application**

- Collecte de retours (feedbacks) d'utilisateurs
  - Avis de divers utilisateurs sur le prototype d'application mis en place sur Hugging Face Spaces, concernant l'interface utilisateur, le confort d'utilisation, l'intuitivité, la qualité des réponses générées
- Amélioration et corrections de l'application et du front-end basé sur le feedback utilisateur
  - Inversion de sens gauche-droite pour l'affichage messages utilisateur-modèle ALOQAS
  - Réduction de taille du logo ALOQAS pour améliorer la visibilité et l'intuitivité de l'interface chatbot avec les suggestions et zone de saisie, sans devoir descendre dans la page
  - Formatage du texte générée en réponse : suppression du prompt apparaissant dans la réponse, suppression des espaces avant les signes de ponctuation
  - Ajout d'un message de bienvenue par défaut du modèle ALOQAS
  - Correction de l'affichage des logos utilisateur-modèle à côté des messages
  - Réinitialisation de la zone de saisie après envoi d'un prompt au modèle
  - Réorganisation / nettoyage du code Gradio