

Rapport 6

Au cours de cette semaine , notre objectif principal a été d'optimiser les performances de notre modèle, et à cet effet, nous avons entrepris une série d'ajustements visant à améliorer de manière significative la qualité des réponses générées.

Une des modifications majeures a été l'augmentation du nombre d'EPOCHS à 15. Cependant, après une analyse approfondie, il est devenu évident que cette modification n'a pas entraîné de changements substantiels dans les performances du modèle.

Face à cela, on a pris la décision de modifier le preset name du modèle, en optant pour "gpt2_medium_en". Nous avons maintenu le nombre d'époques à 15 dans cette configuration initiale, mais les résultats obtenus n'ont pas été aussi pertinents que prévu. En réponse à cela, nous avons décidé d'aller plus loin en augmentant le nombre d'époques à 20.

Cette dernière modification a marqué un changement des performances du modèle.

Avant:

Modification :

```
# modifier les parametre General hyperparameters
BATCH_SIZE = 32
NUM_BATCHES = 500
EPOCHS = 10 # Can be set to a higher value for better results
MAX_SEQUENCE_LENGTH = 128
MAX_GENERATION_LENGTH = 200
#a modifier
GPT2_PRESET = "gpt2_base_en"

# LoRA-specific hyperparameters
RANK = 4
ALPHA = 32.0
```

Résultat:

Après:

Modification:

```
# modifier les parametre General hyperparameters
BATCH_SIZE = 32
NUM_BATCHES = 500
EPOCHS = 20 # Can be set to a higher value for better results
MAX_SEQUENCE_LENGTH = 128
MAX_GENERATION_LENGTH = 200
#a modifier
GPT2_PRESET = "gpt2_medium_en"

# LoRA-specific hyperparameters
RANK = 4
ALPHA = 32.0
```

Résultat: