

Projet : Fine-tuning de Language Models (LLMs)

Dates et modalités de rendu

- **Date limite de rendu** : Vendredi 19 septembre 2025 à 23h59.
- **Travail en binôme**.
- **Format attendu** : Un fichier unique au format Jupyter Notebook (.ipynb) contenant :
 - Les noms et prénoms des membres du binôme ;
 - Le code utilisé pour le projet (avec les visualisations et les résultats) ;
 - Les analyses écrites au format Markdown.
- **Mode de dépôt** : Le fichier doit être déposé via Moodle ou Google Form (un lien sera communiqué ultérieurement).

Objectif

Le but de ce projet est d'adapter l'un des notebooks (*distilbert-finetuning.ipynb* ou *training-tinyllama.ipynb*) vus lors de la dernière séance à un autre cas d'usage. Vous devez réaliser un fine-tuning d'un LLM en modifiant l'un ou plusieurs des éléments suivants :

- le **modèle** utilisé ;
- le **jeu de données** (dataset) ;
- ou les deux.

Cas 1 — Vous modifiez le modèle

- Décrivez le modèle choisi de manière précise et scientifique : nombre de paramètres, architecture, stratégie d'entraînement, etc.
- Comparez-le avec le modèle utilisé dans le notebook d'origine : en termes de performances, d'architecture, etc.

Cas 2 — Vous modifiez le dataset

- Décrivez précisément le dataset :
 - Questions classiques de machine learning : nombre d'exemples, équilibre des classes, etc.
 - Spécificités NLP : longueur moyenne des textes (en nombre de mots), langue(s), etc.
- Analysez les performances du modèle en faisant varier certains aspects du dataset : nombre d'exemples, longueur des textes, preprocessing, etc.

Cas 3 — Vous modifiez le modèle et le dataset

- Faites la description du modèle (premier item du Cas 1).
- Faites la description du dataset (premier item du Cas 2).
- Proposez une petite analyse descriptive des performances obtenues.