**Report – Large Language Models Lab**

# Adaptation de Modèles de Langage Pré-entraînés pour la Classification de Sentiments

**Wassim Chikhi**

*Master 2 Vision et Machines Intelligentes – 2025/2026*

*La soumission comprend un notebook Jupyter (code, résultats, sorties console) ainsi que ce rapport LaTeX.*

**Notebook :** https://colab.research.google.com/
**Auteur :** Wassim Chikhi

## 1. Introduction

Ce TP explore l'adaptation d'un modèle de langage pré-entraîné à une tâche de **classification de sentiments** en comparant trois stratégies : (i) inférence sans entraînement, (ii) *linear probing*, (iii) *fine-tuning* (complet et partiel).

Les expériences utilisent :

— **Dataset :** `cornell-movie-review-data/rotten_tomatoes`
— **Modèle :** `cardiffnlp/twitter-roberta-base-sentiment-latest`

## 2. Données

### 2.1. Dataset Rotten Tomatoes

Le dataset contient des critiques de films annotées en binaire : *NEGATIVE* (0) et *POSITIVE* (1). Le notebook montre la structure : `train` (8530), `validation` (1066), `test` (1066).

| Split | Total | NEG (0) | POS (1) |
|---|---|---|---|
| Train | 8530 | 4265 | 4265 |
| Validation | 1066 | 533 | 533 |
| Test | 1066 | 533 | 533 |

**TABLE 1.** Répartition du dataset Rotten Tomatoes.

### 2.2. Remarque sur le token Hugging Face

L'accès au dataset et au modèle est public et ne nécessite généralement pas de token. Cependant, en environnement restreint, une authentification peut être requise. **Aucun token n'apparaît dans le notebook ou le rapport rendu.**

## 3. Modèle et Tokenisation

Le modèle `twitter-roberta-base-sentiment-latest` est une variante de RoBERTa orientée sentiment, initialement entraînée sur trois classes (*negative*, *neutral*, *positive*).

La tâche cible étant binaire, la tête de classification est réinitialisée en deux classes lors des phases d'entraînement, via l'option `ignore_mismatched_sizes=True`.

## 4. Partie 1 : Inférence sans entraînement

### 4.1. Méthode

Le modèle est utilisé sans entraînement via `pipeline(sentiment-analysis)`. Un mapping binaire est appliqué : *positive* → POSITIVE, *negative* et *neutral* → NEGATIVE.

### 4.2. Résultats

| Classe | Precision | Recall | F1 | Support |
|---|---|---|---|---|
| NEGATIVE | 0.6821 | 0.9418 | 0.7912 | 533 |
| POSITIVE | 0.9061 | 0.5610 | 0.6929 | 533 |
| Accuracy | | | 0.7514 | |

**TABLE 2.** Résultats de l'inférence sans entraînement sur le jeu de test.

## 5. Partie 2 : Linear Probing

Toutes les couches du backbone sont gelées, seule la tête de classification est entraînée (0.475% des paramètres).

| Méthode | Accuracy | F1 | Eval loss |
|---|---|---|---|
| Linear probing | 0.8358 | 0.8351 | 0.3785 |

**TABLE 3.** Résultats du linear probing sur le jeu de test.

## 6. Partie 3 : Fine-tuning

### 6.1. Fine-tuning complet

| Méthode | Accuracy | F1 | Eval loss |
|---|---|---|---|
| Fine-tuning complet | 0.8771 | 0.8744 | 0.5696 |

**TABLE 4.** Résultats du fine-tuning complet.

### 6.2. Fine-tuning partiel

Les deux premières couches de l'encodeur sont gelées (77.25% des paramètres entraînables).

| Méthode | Accuracy | F1 | Eval loss |
|---|---|---|---|
| Fine-tuning partiel | 0.8752 | 0.8725 | 0.5418 |

**TABLE 5.** Résultats du fine-tuning partiel.

## 7. Comparaison globale

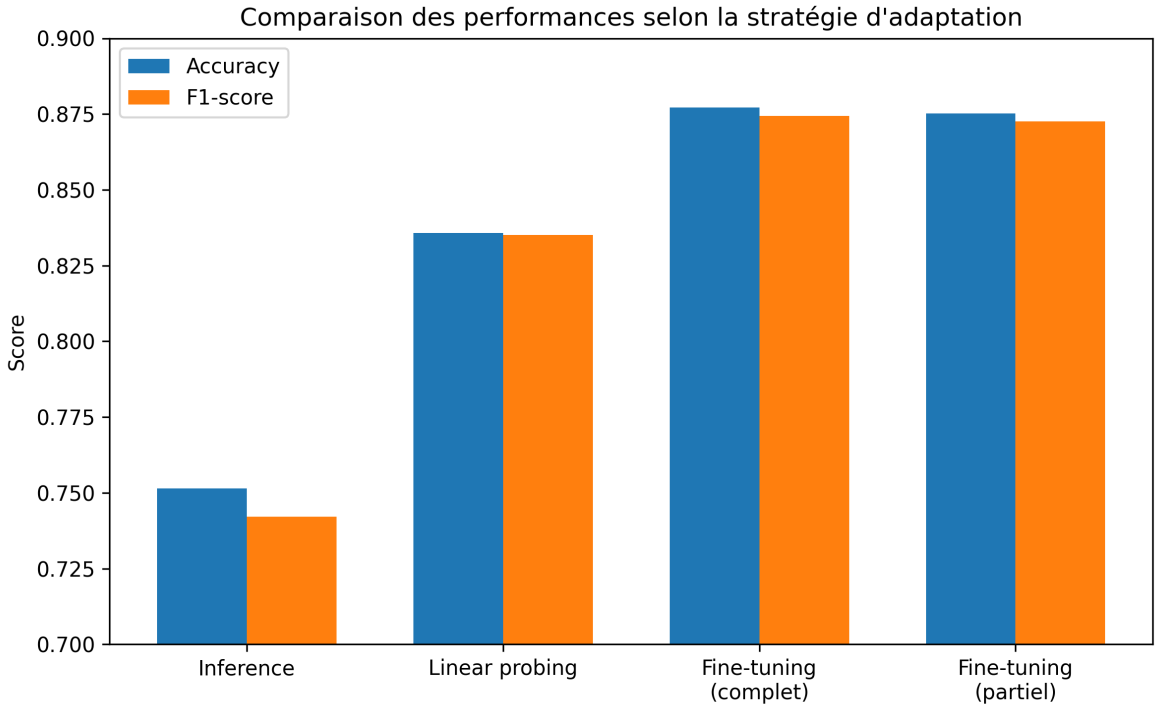


**FIGURE 1.** Comparaison des performances (Accuracy et F1-score) pour les différentes stratégies d'adaptation.

## 8. Analyse détaillée

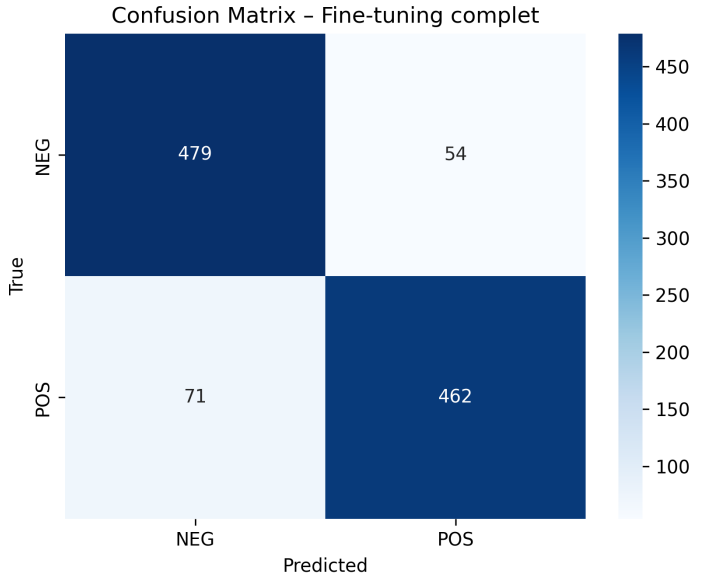


**FIGURE 2.** Matrice de confusion pour le fine-tuning complet sur le jeu de test.

La majorité des erreurs correspond à des critiques ambiguës, tandis que le modèle présente un bon équilibre entre vrais positifs et vrais négatifs, confirmant les performances élevées observées.

## 9. Discussion

Les résultats montrent une progression cohérente des performances lorsque le degré d'adaptation augmente. Le linear probing offre un compromis intéressant entre performance et coût. Le fine-tuning partiel atteint des performances très proches du fine-tuning complet avec un nombre réduit de paramètres entraînables.

## 10. Conclusion

Ce TP met en évidence les compromis entre performance et efficacité lors de l'adaptation des LLMs. Le fine-tuning partiel apparaît comme un choix pertinent dans des contextes à ressources limitées, tout en conservant des performances proches de l'optimum.

**Auteur :** Wassim Chikhi **Année académique :** 2025/2026