# EFFICIENT FEW-SHOT LEARNING WITHOUT PROMPTS

Lewis Tunstall , Nils Reimers , Unso Eun Seo Jo , Luke Bates , Daniel Korat , Moshe Wasserblat , Oren Pereg

# INTRODUCTION

"Efficient Few-Shot Learning Without Prompts" introduces SETFIT (Sentence Transformer Fine-Tuning), a novel approach that simplifies the process. SETFIT eliminates the need for manually crafted prompts, making the few-shot learning process more straightforward and efficient.

## 01 - FEW-SHOT LEARNING

## 02 - SETFIT ELIMINATES THE NEED FOR MANUALLY CRAFTED PROMPTS

1https://github.com/huggingface/setfit
2https://huggingface.co/setfit

# BACKGROUND

01 - FEW-SHOT LEARNING IS CRUCIAL FOR SCENARIOS WITH LIMITED LABELED DATA.

02 - TRADITIONAL METHODS OFTEN FAIL DUE TO RELIANCE ON LARGE DATASETS.

03 - EXISTING METHODS LIKE PEFT AND PET ARE EFFECTIVE BUT HAVE DRAWBACKS:

- High variability due to manual prompts.
- Need for large-scale language models.

N[5]ils Reimers and Iryna Gurevych. 2019. SentenceBERT: Sentence embeddings using Siamese BERTnetworks. I

# SETFIT METHOD

## FRAMEWORK

SETFIT fine-tunes a pre-trained Sentence Transformer (ST) model on a small number of text pairs in a contrastive Siamese manner.

## TRAINING STEPS

1. Fine-tuning ST: Uses positive and negative pairs of sentences to learn meaningful sentence embeddings.
2. Training Classifier: These embeddings train a classification head (e.g., logistic regression) for final classification.

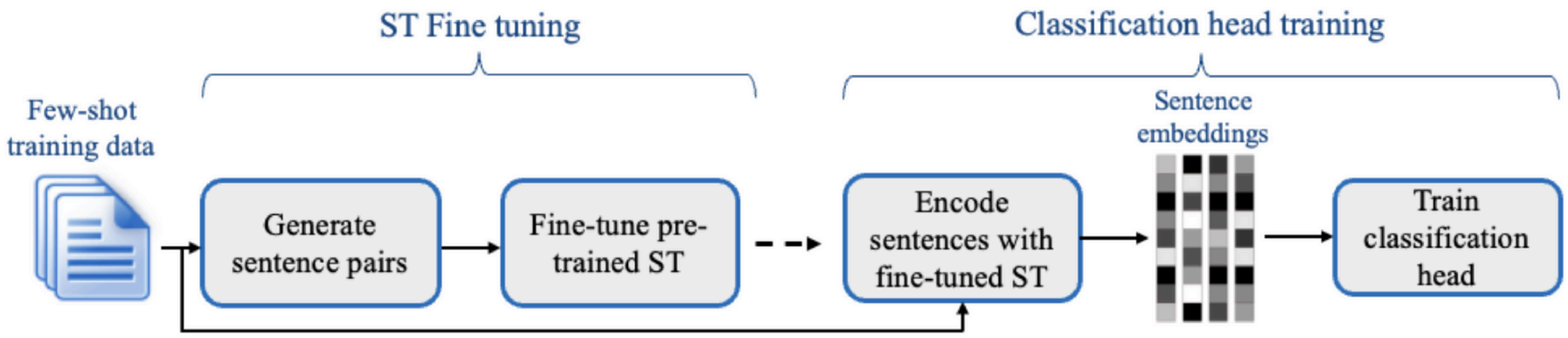



Figure 2: SETFIT 's fine-tuning and training block diagram.

# ADVANTAGES OF SETFIT

- **PROMPT-FREE**
- **REQUIRES FEWER PARAMETERS**
- **MORE EFFICIENT COMPARED TO EXISTING METHODS**

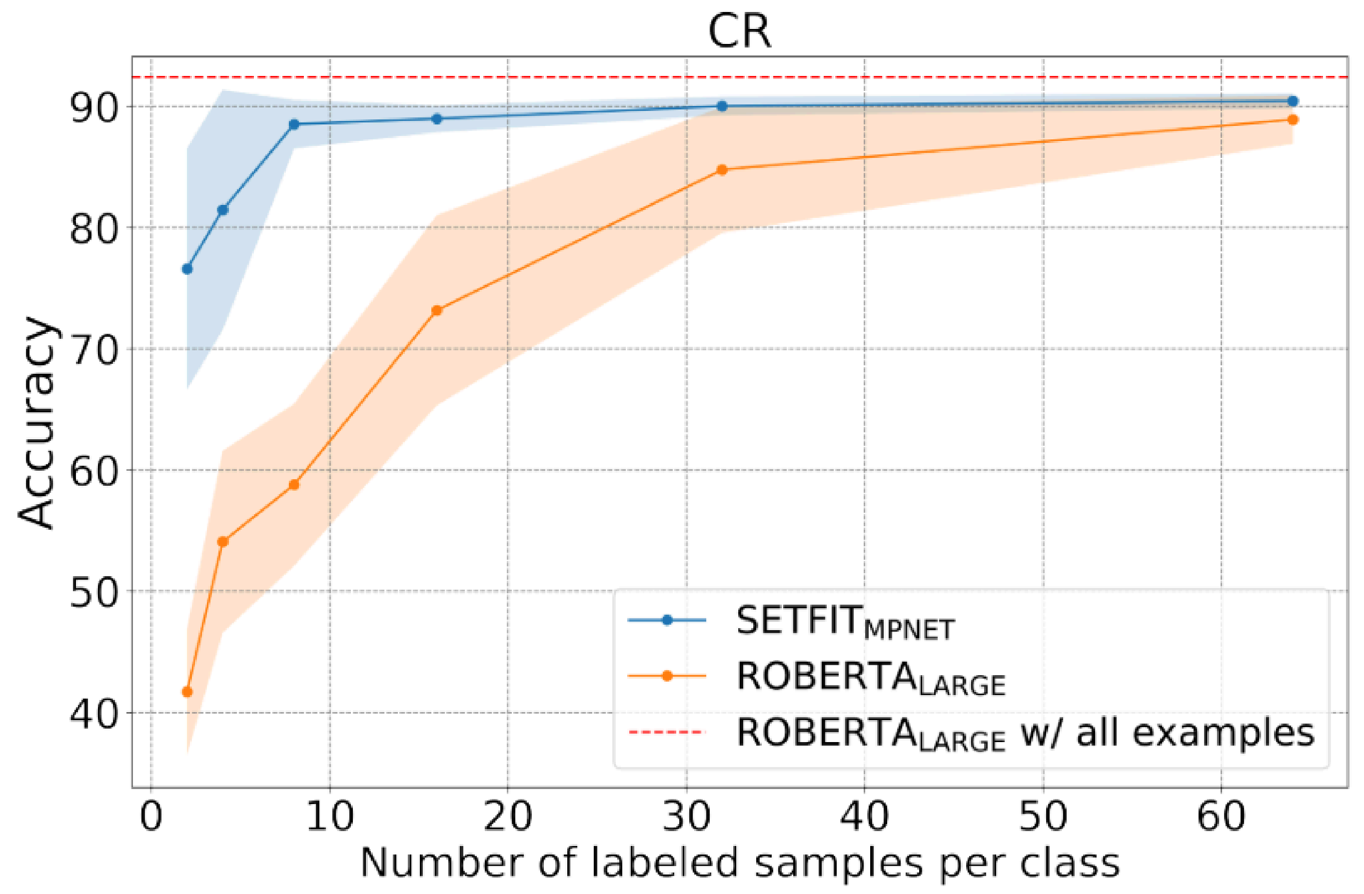


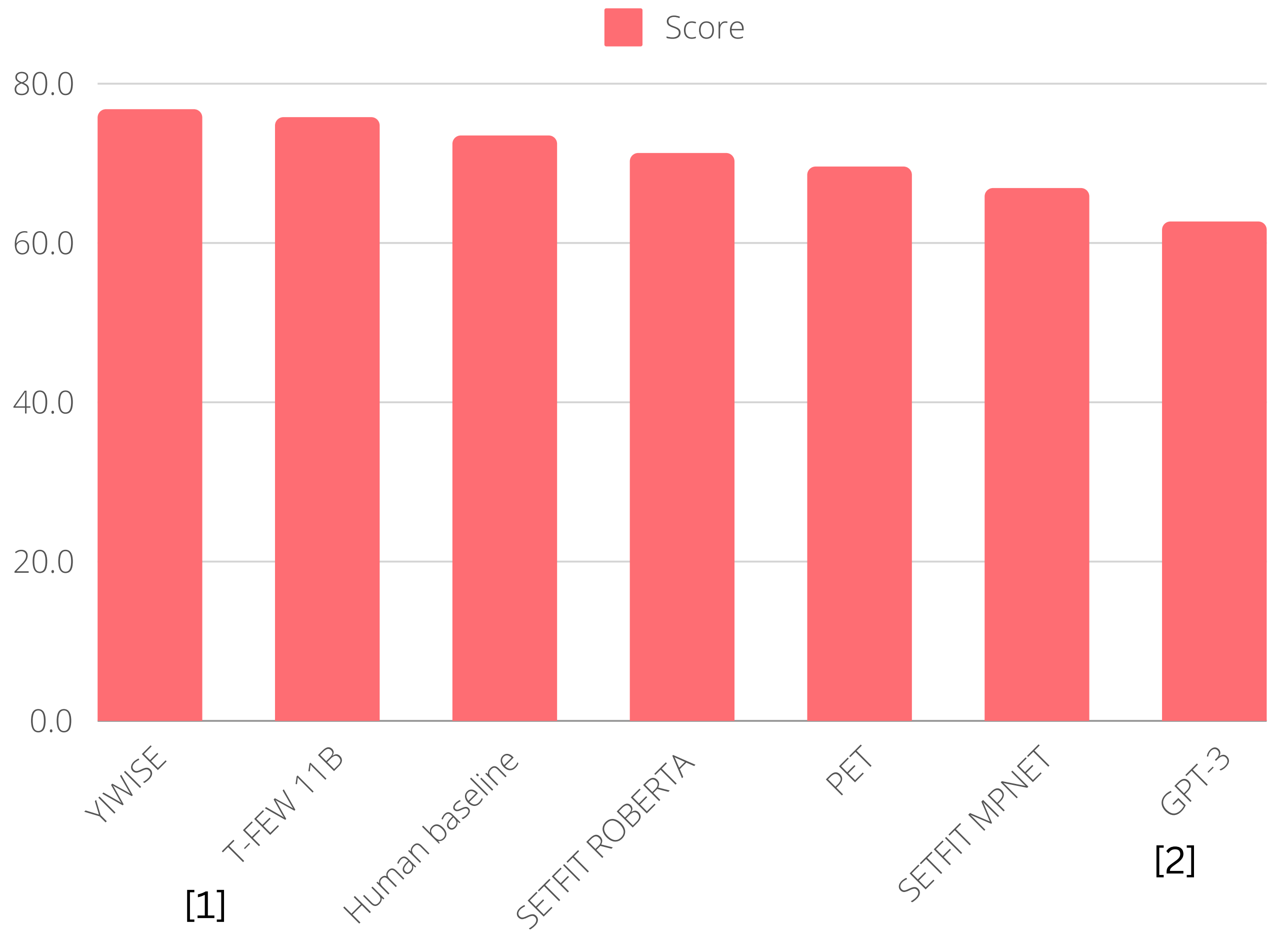

Figure 1: Compared to standard fine-tuning, SETFIT is more sample efficient and exhibits less variability when trained on a small number of labeled examples.

# EXPERIMENTS AND RESULTS

**Performance:**
SETFIT achieves comparable results to state-of-the-art few-shot learning methods while being faster and more efficient.

**Datasets:**
The method was evaluated on various text classification tasks, demonstrating its robustness across different types of data and multilingual scenarios.

**Efficiency:**
SETFIT models are smaller and require less computational power for training and inference, making them suitable for real-world applications.

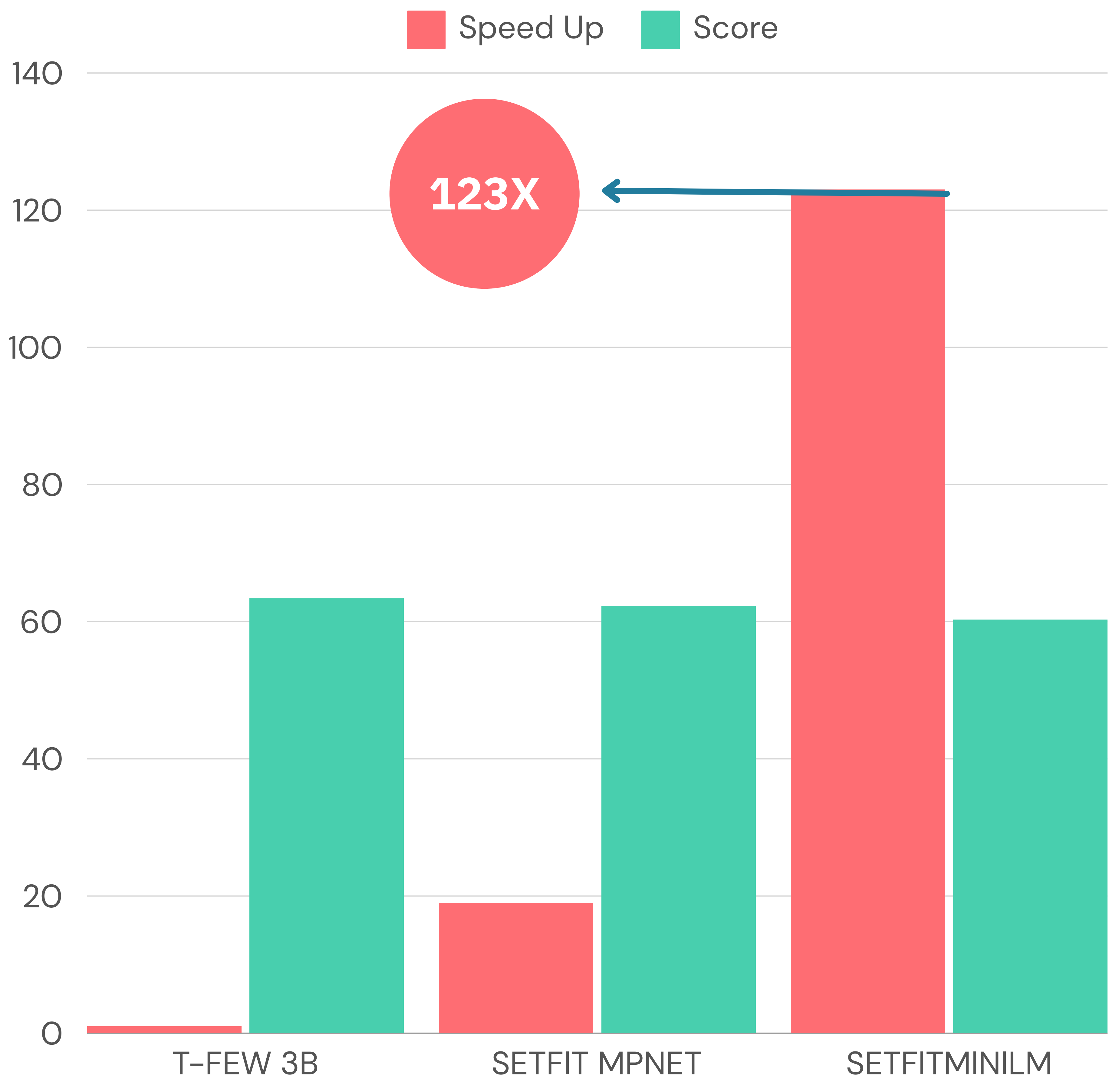



[1]Haokun Liu, Derek Tam, Mohammed Muqeeth, Jay Mohta, Tenghao Huang, Mohit Bansal, and Colin Raffel. 2022. Few-shot parameter-efficient finetuning is better and cheaper than in-context learning.

[2]Tom Brown, Language models are few-shot learners. In Advances in Neural Information Processing Systems, volume 33, pages 1877–1901. Curran Associates, Inc.

# COMPUTATIONAL COSTS

- **Training and Inference Efficiency**
- **FLOPs Comparison[3]**
- **Model Size**
- **Real-world Deployment:**
- **Training Time**




[3]Jared Kaplan, Sam McCandlish, Tom Henighan, Amodei. 2020. Scaling laws for neural language models. CoRR, abs/2001.08361.

# MULTILINGUAL CAPABILITIES

## 01 - Multilingual Training

SETFIT can be easily adapted to different languages by switching the underlying Sentence Transformer model.

## 02 - Performance

It shows strong performance in few-shot learning scenarios across multiple languages, outperforming traditional fine-tuning and ADAPET (an advanced few-shot learning method)[4]

| Method | Train | En | De | Ja | Zh | Fr | Es | Average |
|---|---|---|---|---|---|---|---|---|
| | | | | $\lvert N\rvert = 8^*$ | | | | |
| FINETUNE | each | $122.9_{14.0}$ | $119.9_{13.6}$ | $120.5_{8.0}$ | $128.6_{10.7}$ | $123.2_{13.0}$ | $116.3_{8.3}$ | $121.9_{11.3}$ |
| | en | $115.9_{11.3}$ | $115.2_{12.0}$ | $121.6_{12.3}$ | $123.0_{8.8}$ | $117.3_{13.0}$ | $113.1_{12.4}$ | $117.7_{11.6}$ |
| | all | $117.8_{4.9}$ | $116.3_{9.7}$ | $121.5_{12.4}$ | $120.5_{6.7}$ | $117.3_{9.9}$ | $110.1_{9.5}$ | $117.2_{8.8}$ |
| ADAPET | each | $129.9_{13.6}$ | $136.4_{10.6}$ | $130.4_{13.4}$ | $135.0_{10.9}$ | $141.8_{10.1}$ | $136.0_{10.4}$ | $134.9_{11.5}$ |
| | en | $138.9_{17.8}$ | $151.5_{17.8}$ | $160.8_{16.7}$ | $158.8_{16.3}$ | $152.0_{15.7}$ | $149.8_{17.1}$ | $152.0_{16.9}$ |
| | all | $150.8_{12.0}$ | $136.2_{7.0}$ | $150.8_{10.0}$ | $152.8_{10.2}$ | $140.0_{14.0}$ | $145.1_{4.5}$ | $146.0_{11.3}$ |
| SETFIT | each | $82.9_{4.3}$ | $\mathbf{80.0}_{2.4}$ | $95.5_{2.8}$ | $95.3_{2.8}$ | $85.3_{6.0}$ | $\mathbf{80.8}_{5.4}$ | $86.6_{4.9}$ |
| | en | $\mathbf{82.6}_{4.8}$ | $83.4_{5.9}$ | $\mathbf{93.2}_{6.6}$ | $\mathbf{93.9}_{3.6}$ | $\mathbf{82.2}_{4.8}$ | $83.4_{5.9}$ | $\mathbf{86.4}_{5.2}$ |
| | all | $83.0_{5.3}$ | $84.0_{7.6}$ | $97.1_{9.2}$ | $97.4_{6.5}$ | $83.5_{6.5}$ | $84.9_{6.1}$ | $88.3_{6.9}$ |
| | | | | $\lvert N\rvert = Full^{**}$ | | | | |
| FINETUNE | each | 46.2 | **43.7** | **46.8** | **56.6** | **47.8** | **45.3** | **47.7** |
| | en | **46.1** | 46.6 | 61.0 | 69.4 | 55.6 | 52.9 | 55.3 |
| | all | 46.6 | 49.4 | 61.0 | 69.4 | 55.6 | 55.0 | 56.2 |

Table 4: Average performance (MAE $\times$ 100) on the Multilingual Amazon Reviews Corpus for two training set sizes $\lvert N\rvert$. $^*$ No. of training samples per class. $^{**}$Entire available training data used (20,000 samples).

[4]Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale. CoRR, abs/1911.02116.

# CONCLUSION

SETFIT is a significant advancement in few-shot learning, offering a practical and efficient alternative to existing methods. Its simplicity, efficiency, and prompt-free nature make it an attractive option for scenarios with limited labeled data.

# DEMO

# THANK YOU