



Digital Research Alliance of Canada: Project Manuals, Tutorials, and User Guides

Tutorial: Building a Fake News Classifier on HPC

Objective: Train a transformer-based model for fake news detection on a supercomputer.

Sections:

1. Choosing a Dataset (LIAR, FA-KES, Snopes)
2. Preprocessing for Text Classification
3. Using HuggingFace Transformers with PyTorch
4. Hyperparameter Tuning in Parallel
5. Saving Models and Checkpoints to Project Space
6. Post-training Evaluation and Export
- 7.

```
from transformers import BertForSequenceClassification, Trainer
model = BertForSequenceClassification.from_pretrained("bert-base-uncased")
# ... Trainer config ...
trainer.train()
```