

Appendix C. Bert fine-tuning parameters

Category	Parameter	Value
Architecture	Base model	google-bert/bert-base-uncased
Batch sizes	Train batch size	8
	Eval batch size	16
Learning rate	Initial learning rate	5E-05
Scheduler	Type	Linear
	Warmup ratio	0.1
Regularization	Weight decay	0.0
	Max gradient norm	1.0
	Dropout rate within each attention head	0.1
	Dropout rate in hidden layers	0.1
Training duration	Epochs	3
Evaluation	Eval strategy	Per epoch
	Save strategy	Per epoch
Loss function	Type	Cross-entropy
Randomness control	Seed	42
Optimiser	Type	AdamW (adamw_torch), $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1\text{e-}8$

Note. All models were trained with identical hyperparameters; all other hyperparameters were default but can be accessed in the GitHub repository under `.../02_code/00_setup_requirements/BERT_fine_tune_args.py`. Initial experiments with different training durations showed that the models learned quickly, with little improvement after early epochs. To reduce the risk of overfitting, fine-tuning was fixed to three epochs. At the end of training, the checkpoint with the lowest validation loss (evaluated once per epoch) was reloaded for subsequent evaluation and prediction.