



# BOOLQ CLASSIFICATION

---

Yelin Zhang

# DECISIONS

- Preprocessing
- Model
- Training

# PREPROCESSING

- Lower case before tokenization
- Lemmatization
- Fasttext
- Truncating long passages

# MODEL

- Input: Concatenation of padded question and passage
- Dimensional reduction (`torch.squeeze`)
- Optimizers: Adam or AdamW
- Loss: Binary Cross Entropy or HingeEmbedding

# TRAINING

- Enable dataloader batching and accelerator
- Optuna hyperparameter sweeping

# SHOW & TELL

- ruff

```
classifier.ipynb:cell 14:30:5: F821 Undefined name `wand`  
28 |  
29 |     )  
30 |     trial.set_user_attr(key="model", value=classifier)  
    |     wand.finish()  
    |     ^^^^ F821  
31 |  
32 |     return trainer.callback_metrics["train_accuracy"].item()  
    |  
Found 1 error.
```

# REFLECTION

- Clarification of task before decisions
- More parameters instead of dimension reduction
- Use wandb built in sweep function (better integration)