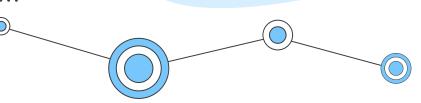# Week 4: Labeling Data Pt. 2

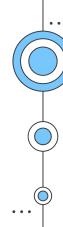Alfonso Vieyra, Denise Wei, Raha Ghazi, Tom Gao

Labeling with Flan-T5-Large

- Issue with token sequence length (512 max)
  - Using longformer for the longer dialogue?
  - Currently rows of longer dialogue are unlabelled
- Still about 80% positive, 20% negative
  - Plan on undersampling positive rows
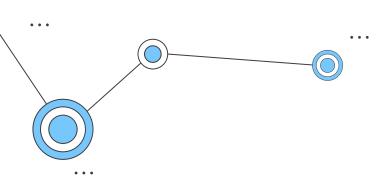- Very slow process, was broken up in batches

Accuracy?

- Over 90% accurate when manually assessed over 200 rows
- Better ways to determine accuracy?
- Value of labelling with other models and comparing results?

Target LLMs to Evaluate

- Mistral
- Llama 2
- BERT
- BLOOM
- GPT-J
- MPT
- Falcon
- XGen
- Zephyr
- Dolly