19/02 Meeting

Points to go over:

* Went through dataset once removing irrelevant content in 20 most common trigrams, 4-grams, 5-grams for each outlet.
* Tested baselines again, not a very noticeable difference, tried tests like using a smaller sample of the dataset(0.5,0.3), cutting each article down, removing common keywords, baselines either dropped slightly or actually improved in some cases
* Went through dataset again, removing any duplicate articles that evaded me, and removing irrelevant content in 40 most common trigrams, 4-grams,5-grams for each outlet.
* Tested baselines with different tests, still very high results(show excel file).
* Do I accept that the models are good and focus on different tests I can do or do I continue trying to find artifacts in the dataset? What are good tests I can do or what are methods I can use to find artifacts?

Minutes:

* Think I’m fine to accept current dataset and move on to making different tests
* Could do sentence-level testing
  + Determine which sentences get high scores
  + Try and get visualisations
  + Can do that through spacy(sentence-splitting, look for parser token flags)
  + Works for both baselines and fine-tuned LLMs
  + Multiple-instance learning, dealing with bags of samples instead of one sample at a time. Don’t look into that too much, a whole different research project, a good note for the diss.