Grad research outline

* Intro
  + What is the problem?
    - LLMs are better than ever and can be used to generate content that is eerily like what humans create.
    - This is a problem for academia as students can use this technology to cheat.
    - Some schools are adapting, but others are relying on technology to detect ai content.
    - These tools don’t work well.
  + What is my solution?
    - I don’t have one, this is an investigation into the characteristics of AI content and how it can potentially be detected.
    - Generated lots of text based on many topics. Created several models based on those topics.
  + Why should we care?
    - This is going to continue to be a problem.
    - Existing tools do not work well.
* Literature
  + What/when/how is LLM?
    - LLM – Large Language Model, model trained on massive amounts of text data, close to the entire internet is being used as dataset.
    - These models have been around since the 1950’s, in theory, but were impractical until 2017
  + “Attention is All You Need”
    - Research paper by google, detailed the architecture of the Attention model. Unlike RNN and LSTM, these models were easily trained across many computers simultaneously.
    - Training went from taking months/years, to weeks.
  + ChatGPT/OpenAI
    - OpenAI is the company, ChatGPT is the public facing product, GPT 3.5/4 is the model.
    - Fasting growing internet product in human history, 100 million users in first 4ish months.
    - Microsoft is the primary investor.
  + Detection tools
* Experiments
  + Set up/hypotheses
    - Explore whether we can find out some of why models have trouble with this
    - Explore how openai’s fine tuning models work, and how it performs on NLP tasks
    - Compare texts of different grade levels/quality
    - Compare texts of human vs ai
  + Methodology
    - Gather human created texts
      * Medium.com dataset, use article titles and texts
    - Gather ai text
      * Use medium titles to generate texts
      * Use openai API to systematically generate those texts
    - Create dataset for classification
      * Get all ai data into dataframe with labels
      * Get all ai vs human data into dataframe with labels
    - Upload to finetuning
      * Fine tuning does it’s own preprocessing, and recommends which pre-trained model to select.
      * Do classification
    - Download results
    - Create tensorflow model using same dataset
      * Create functions to calculate metrics
        + Average sentence length
        + Average sentence complexity
        + Sentiment score
    - Run LSTM model in tensorflow
      * 3 class all ai
      * Human vs ai
  + Results
  + Interpretation
* Future work
* Bibliography