**Meeting Minutes**

*Attendees:*

Dr. Diego Molla-Aliod

Nour Chalouhi

**Date / Time: 4PM 12/03/2024**

**Discussions:**

* Key points discussed:
  + Attributes of machine learning pertinent to the project.
  + Techniques in Summarisation
  + Framework for system implementation
  + Scoring System
  + Common Techniques for prompt Engineering
  + Fact Checking and AI Hallucination mitigation.
  + criteria for assessing summarisation quality.
  + Error analysis
  + Improving quality of summarisation
  + Central issues within the scope of this research field.

**To Do:**

* Review literature and databases emailed by supervisor:

1. **Readability Controllable Biomedical Document Summarisation** by Zheheng Luo, et al.: Introduces a corpus with 28,124 biomedical papers for readability-adjustable summarisation, showcasing transformer-based methods. [Paper](https://arxiv.org/abs/2210.04705) | [Corpus](http://www.nactem.ac.uk/readability/)
2. **Making Science Simple: Corpora for the Lay Summarisation of Scientific Literature** by Tomas Goldsack, et al.: Offers a general corpus for scientific literature summarisation, not limited to biomedical texts. [Paper](https://arxiv.org/abs/2210.09932) | [GitHub](https://github.com/TGoldsack1/Corpora_for_Lay_Summarisation)
3. **A Dataset for Plain Language Adaptation of Biomedical Abstracts** by Kush Attal, et al.: Provides a dataset of 750 abstracts and 7,653 sentence pairs for biomedical abstracts' plain language adaptation. [Paper](https://arxiv.org/abs/2210.12242) | [Dataset](https://osf.io/rnpmf/)

**Completed:**

* Reviewed academic papers.
* Analysed databases for structural insights of biomedical papers.

**Date / Time: 11:30AM 19/03/2024**

**Discussions:**

* Shared insights from literature on evaluation metrics, with a focus on ROUGE.
* Highlighted the importance of choosing the most relevant evaluation frameworks for summarisation output assessment.
* Discussed the use of various AI platforms for preliminary summarisation experimentation, including Chat GPT, Gemini, and PaLM 2, deciding on Gemini for its cost advantage.
* Recommended to download the ROUGE evaluation toolkit via the Anaconda platform.
* Emphasised finding the most relevant evaluation metric tailored to the objectives of the study.

**To Do:**

* Advised to write some code to summarise biomedical papers and compare to the already summarised paragraphs.
* Advised to try some prompts using the PLOS sample.
* Emailed by Supervisor:

*" I just read the Google's PaLM is no longer supported, instead you can use Gemini. You can read the documentation here:*[*https://ai.google.dev/docs*](https://url.au.m.mimecastprotect.com/s/RgsPCvl1g2Smw1J6SXa7R1?domain=aus01.safelinks.protection.outlook.com)

*You have the choice to obtain a developer key and develop the code in your computer (the preferred choice), or develop using Google AI Studio (which may be good for the first time and for quick tests)"*

* Create a Git Repository to add meeting minutes and all related documentation.

**Completed:**

* Downloaded Git.
* Set up development environment by downloading Python and the Anaconda distribution.
* Switched back to using Visual Studio Code as my primary IDE.
* Installed the conda package manager, Python extensions for IDE, and other relevant packages necessary for my project.
* Wrote a Python script that uses the integrated API 'Gemini Pro' from Google's Generative AI to automatically summarise biomedical research papers.
  + The script reads from a JSONL file containing document abstracts and sequentially generates summaries, restricting them to the first 200 words
  + Each summary is numbered and formatted for clear differentiation, with error handling in place to manage any potential issues during the summarisation process.

**Date / Time: 11:30AM 26/03/2024**

**Discussions:**

* Shared summary generator created on VsCode, integrated with the Gemini API
* Successfully executed summarisation of 1000 biomedical abstracts using Gemini.
* Addressed the limitations and scope of using ChatGPT in the project.
* Discussed challenges to be considered associated with the thesis topic.
  + There are numerous summary techniques including abstractive and extractive methods.
  + Scoring metric are subject to change in response to the unpredictable nature of AI-generated text.

**To do:**

* To integrate the ROUGE Metric System into the Python program for the evaluating summarisation output.
* To compare the plain language summaries against AI-generated summaries and produce overall ROUGE scores.
* To continue with a literature review.
* To compile the developed program and meeting minutes into a Git repository and share with Supervisor.

**Completed:**

* Downloaded ROUGE
* Incorporated the ROUGE Metric System.
* Downloaded Zotera for effective Citation Management
* Literature Review

**~~Date / Time: 4:30PM 02/04/2024~~**

* Cancelled and Rescheduled to 09/04/2024

**Date / Time: 11:30AM 09/04/2024**

**Discussions:**

* Discussed a publication titled "Control Stochastic Selection-Based Biomedical Text Summarization Using Sim-TLBO", however, I was advised to concentrate efforts on prompt engineering.
* Introduced to the resource "Paperswithcode", which provides access to scientific papers along with their corresponding code repositories and datasets.
* Agreed to prioritise research and application of prompt engineering methods in the context of the project.

**To do:**

* To research further on the methodologies and best practices of prompt engineering.
* To continue with literature review
* To continue with code experimentation and testing.
* Begin my thesis writing.

**Completed:**

* Attempted to continue testing however I need to find a quicker way to complete testing.
  + Finding difficulty in data handling
  + Also, Characters being returned from my output are limited to 100.
  + Need to find a better way of processing data.
* Notes to self:
  + List 5 different extractive ways to identify Prompts vs abstractive and give scoring.
  + **Extractive** is more suitable when you need to ensure the summary is highly accurate with respect to the source content, such as for legal or technical documents.
    - **Legal implications – less erroneous**
  + **Abstractive** is better when you want a more readable summary that might be easier to understand and shorter, suitable for general knowledge.
    - **More readable for lay person**
  + How to handle such datasets
  + How many data sets should I do different testing on or should I stay testing on one data set and reference it in my framework.
  + Extractive then abstractive method
  + Meteor scoring – synonym matching – more effective for layperson summary judgements
  + Highlight that can be useful
    - Summarise
  + Look at current abstract – comp
  + Analysis of t
  + Overfitting
  + System while training could because take
* Training data test unseen to look at
* List papers

\