**Meeting Minutes**

*Attendees:*

Dr. Diego Molla-Aliod

Nour Chalouhi

**Week 4**

**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.

**Week 5**

**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.

Week 6

**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

**~~Week 7~~**

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

* Cancelled and Rescheduled to 09/04/2024

**Week 8**

**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**

**Break Week 2**

**Date / Time: 2:00 PM, 23/04/2024**

**Discussions:**

* Handling datasets and testing methodologies.
* Number of datasets for different testing phases.
* Extractive then abstractive method potential.
* Meteor scoring – synonym matching – more effective for layperson summary judgments.
* Summarisation analysis and overfitting considerations.
* Training data with unseen corpus.

**To Do:**

* Address GitHub repository issue raised by the supervisor.

**Completed**:

**Week 9**

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

**Discussions:**

* Structure of the thesis.
* Organised scores into tables.
* Preliminary testing with the use of persona, chain of thought, and few-shot techniques, organising average ROUGE 1,2,L scores into a table.
* Literature review and code refinement.

**To Do:**

**Completed:**

**Date / Time: 4 PM, 07/05/2024**

**Discussions:**

* Penalising jargon in training objectives to improve model outputs, increasing simplicity and abstractedness of summaries.
* Importance of lay summarising.
* Presentation preparation for week 12.
* Advances in the field and primary outcomes of testing.
* Deviation from existing approaches.
* Tangible economic benefits.
* GitHub table and preliminary chain of thought.
* More data and prompt considerations.

**To Do:**

**Completed:**

**Date / Time: 11:30 AM, 14/05/2024**

**Discussions:**

* Methodology and preliminary results.
* Avoiding wordiness in extractive summarisation.
* Adopting a persona for abstractive summarisation.
* Flesch-Kincaid grade level for readability.
* Penalising jargon.
* Comparison prompts.

**To Do:**

* Explore prompt engineering resources:
  + [ChatGPT Prompts to Summarize Text](https://nerdschalk.com/chatgpt-prompts-to-summarize-text/)
  + [ChatGPT Prompts for Summarizing](https://chatgptaihub.com/chatgpt-prompts-for-summarizing/)
  + [Prompt Engineering Guide](https://www.promptingguide.ai/introduction/examples)
  + [Dynamic Summarization Prompts](https://ai.plainenglish.io/10-chatgpt-prompts-for-dynamic-summarization-2441f92fe523)
  + [Prompt Engineering for Summarization](https://www.geeksforgeeks.org/prompt-engineering-for-summarization/)
  + [GPT Summary Prompt Engineering](https://devblogs.microsoft.com/ise/gpt-summary-prompt-engineering/)
  + Relevant papers: [arxiv.org/abs/2304.14670](https://arxiv.org/abs/2304.14670), [arxiv.org/abs/2403.02901](https://arxiv.org/abs/2403.02901)

**Completed:**

* **Reviewed papers.**

**Date / Time: 11:30 AM, 21/05/2024**

**Discussions:**

* Presentation discussion.
* BioLaySumm 2024 dataset for testing.
* Recent competition papers and their prompts.
* Using these prompts or getting ideas from their work.
* Relevant paper: [arxiv.org/abs/2405.11950](https://arxiv.org/abs/2405.11950)
* Paper on RAG and reinforcement learning: [arxiv.org/abs/2405.13179](https://arxiv.org/abs/2405.13179)

**To Do:**

* + Work on thesis
  + Wrap up methodology gaps regarding structure issues etc as the PLS in the corpora does not have a Structured written PLS

**Completed:**

* Completed presentation.
* Discussed improvements and plans for next week to read through my thesis.

**Date / Time: 3:30PM, 31/05/2024**

**Discussions:**

* + Sent Methodology earlier that day, supervisor reviewed
  + Discussed feedback
    - More clarity in methodology
    - Asked to analyse data that will be utilised in Thesis B further to ensure all I’m implementing is correct and to not assume aspects of Abstracts will be there etc.
  + Discussed if it is correct justify using the same BIOLAYSUMM 2024 scoring metrics as it will be better to compare with other projects.

**To Do:**

* + Refine methodology and send entire thesis completed