**Practicum Report**

**Introduction**

Generative models have revolutionized the field of natural language processing, enabling the generation of descriptive captions for images. In this project, we aim to develop a generative model and applicable chatbot of caption using the dataset from WikiQA (https://paperswithcode.com/dataset/wikiqa). This dataset provides a valuable resource for training and evaluating our model's performance in generating informative and contextually relevant captions.

The project overview involves utilizing deep learning algorithms and techniques to train a model on the WikiQA dataset. The dataset comprises a large collection of paragraph-question pairs, providing rich and diverse examples for our model to learn from.

**Previous Work:**

"Asking Questions the Human Way: Scalable Question-Answer Generation from Text Corpus" by Liu et al. introduces a method for generating questions and answers from a text corpus in a scalable manner, providing a valuable approach for automated question generation from large datasets.

"Capturing Greater Context for Question Generation" by Tuan et al. proposes a method to capture a wider context for question generation by incorporating information from surrounding sentences, leading to more comprehensive and contextually relevant questions.

Data Description&Preprocess:

Methodology:

My model architecture consists of two main parts: text summarization and Q&A. In the first part (text summarization), I will be working with T5 from Hugging Face Library. In the second part (Q&A), I will be fine-tuning OpenAI's GPT3 to adapt our dataset using suitable prompt.

In the current stage of NLP, training a model that gives decent results most likely requires a huge amount of data as well as computational resources. Due to the tight schedule and limited amount of accessible resources, it is more realistic for us to focus on fine-tuning pretrained models as the building blocks for our final model.

The reason that I came up with this idea is about scale. Instead of storing entire research articles for the model to remember, I can first apply a text summarization model to extract the key

Metric:

In the project, I employed two metrics, ROUGE and BLEU (F1 score), to evaluate the performance of our text generation model. ROUGE (Recall-Oriented Understudy for Gisting Evaluation) measures the overlap between the generated text and reference summaries by computing the recall of n-gram matches. It considers various factors such as unigram, bigram, and longest common subsequence (LCS) matches to assess the quality of the generated text. On the other hand, BLEU (Bilingual Evaluation Understudy) is a widely used metric that evaluates the similarity between the generated text and reference translations by comparing n-gram precision and incorporating a length penalty. We specifically utilized the F1 score variant of BLEU, which strikes a balance between precision and recall, providing a comprehensive evaluation of the text generation quality in terms of matching the reference text and covering the reference content. By employing these metrics, we aimed to obtain a comprehensive and objective assessment of our text generation model's performance.

Deliverable

The deliverable of my project is a chatbot that has the capability to generate concise, one-sentence captions for inputted paragraphs. Leveraging natural language processing techniques and machine learning algorithms, the chatbot analyzes the input paragraph to extract key information and generate a succinct and informative summary in a single sentence. This chatbot aims to provide users with a quick and efficient way to obtain condensed summaries of longer texts, making it easier to grasp the main points and key takeaways without having to read through the entire paragraph. The generated captions aim to capture the essence of the inputted content, enabling users to obtain quick insights and make informed decisions based on the summarized information.