Original Proposal

1. Analyze existing works on query-focused summarization. There are many datasets that are suitable for this task, such as **AQuaMuSe** and **HowSumm** which are both dedicated to query-focused summarization. In addition there are also many models that perform well on summarization tasks, such as **PEGASUS** and **BART**. Finally we need some evaluation metrics to measure how well a model performs. Common metrics on this task includes ROUGE but there are some improvements to this metric we can use such as **ParaEval** and **ROUGE 2.0**. There is doubt on summarization metrics that they do not correlate well with human judgement[1], so we can also try to use surveys to collect human feedback. For example, one approach is to present users with model generated answer and the reference answer and ask them which one they prefer[6].
2. Reproduce existing works as well as testing multiple retrieval methods on this task. Many retrieval methods had good results when integrated with language models, such as **kNN**, **DPR**, **TF-IDF** and **BM-25**. Going beyond retrieval methods we can also focus on directly handling long inputs. For example several works proposed linear time attention mechanism to allow transformers taking longer inputs, such as **LED**, **Big Bird** and **LongT5**.
3. Propose new retrieval-based methods to improve the model and build a model that can do both retrieve and summarize if there's time.

Summer Plan

4. Gather background information on retrieval-augmented language models (LMs) and query-focused summarization. Retrieval LMs are gaining traction due to their increased parameter efficiency and improved interpretability. Several approaches have been proposed, such as utilizing a kNN datastore, constructing specialized structures, and integrating them into models for generation tasks. For knowledge-intensive tasks, methods like Dense Retriever (DPR) have been used to retrieve full documents. The retriever can be applied at various stages, such as pre-training and fine-tuning, and optimized using different objectives. My initial goal is to investigate the range of techniques used thus far and compare their strengths and weaknesses.
5. Replicate these techniques on query-focused summarization datasets. I plan to implement these techniques on some top-performing models to evaluate their performance and analyze factors such as retrieval relevance and computational efficiency. For retrievers that retrieve documents, I will also assess the retrieval quality using only the retriever to measure stand-alone performances.

References:

Exploring Neural Models for Query-Focused Summarization.
- This paper does a similar study where the authors found models that can do the task and used them on query-focused datasets.

Atlas: Few-shot Learning with Retrieval Augmented Language Models
- This paper proposes the model *Atlas* that is designed by rigorously considering different retrieval architectures and approaches.