CSE 156 Proposal

Andrew Pegg, Jason Boenjamin, Ryan Livengood, Christopher Schrader
apegg@ucsd.edu   jboenjamin@ucsd.edu   rlivengood@ucsd.edu   cwschrader@ucsd.edu

Abstract

The rapid growth of scientific research over last couple of decades has created a new difficultly for researchers and the general public to stay informed about developments and achievements across fields of research. Understanding research has also become a challenge for many, as much of today's research is highly esoteric and inaccessible to be understood by mass audiences. Our project aims to address this issue by providing an accessible service that allows users to upload scientific research, and have an AI model summaries the content in understandable and simplified terms. We plan to achieve this through applying several natural language processing (NLP) techniques/processes to further fine tune existing LLM models into the perfect scientific summarizer.

1 Start Up Idea

TLDRsrch - Using LLM to target PhDs / research paper-related queries.

Academic researchers spend countless hours sifting through verbose research papers to extract relevant insights. This inefficiency reduces the amount of time that can be put towards new research and findings and also restricts access to cutting-edge research. By leveraging NLP techniques such as (?) Byte-pair encoding, Wordcloud, citation analysis, and summarization, our model simplifies the complex topics in research papers while maintaining scientific rigor.

Our target users are university students, people who are trying to get into research without formal experience, and industry professionals who rely on cutting-edge research. This solution is valuable because it reduces time spent sifting through literature, and democratizes access to new and exciting scientific discoveries.

2 Market and Competitive analysis

We identified one competitor in the market and analyzed the approaches that they made publicly known. They offer the same product we offer, summarizing research papers, but we believe our product will differ in several ways. One competitor, SciSummary, SciSummary [2025], uses GPT-3.5 and GPT-4 under the hood with human-guided training from PhD holders. One paint point to notice here is that humans may have implicit bias about a particular research paper, thus adding bias to their model.

Our product is similar to this competitor in the sense that we are both using a fine-tuned LLM. However, our product is distinct because we will be fine-tuning our model based on datasets rather than human-guided training. Additionally, our model will be a fine-tuned version of DeepSeek, which we believe to be an improvement on the GPT models our competitor fine-tuned. In summary, we believe our model will be less computationally expensive due to the use of DeepSeek and have less potential for human bias.

3 Technical Approach

Our technical approach for TLDResearch involves fine-tuning a LLM (DeepSeek) with loRA to efficiently summarize research papers while keeping key scientific insights. We will also be creating our own preprocessed pipeline to clean, segment, and structure research papers before feeding them into the model. This will ensure more accurate and context-aware summaries. Additionally, we intend on extracting key features from the research paper and providing that context to our LLM so it can effectively summarize. Another main component of our approach is data engineering with the way we parse PDFs, which includes extracting text, handling figures, and preserving document structure for more coherent outputs. A key technical challenge is ensuring that our model is useful on a wide variety of research papers. We intend on addressing this issue with a proper dataset that includes a variety of research topics. If some varieties are underrepresented, we can apply oversampling.

Our model will be fine-tuned to effectively summarize research papers and provide proper citations. The datasets we will be using to fine-tune our model with this in consideration can be found here respectively: scitldr and s2orc.

4 Implementation plan

The MVP for TLDResearch is a seamless research paper summarization tool that allows users to upload a PDF, process its contents using NLP models, and generate an easy-to-digest summary. The core features that define our product include automatic summarization, citation context extraction, and a simple UI where users can drag and drop PDFs and provide feedback through a thumbs up and thumbs down button. The thumbs-down button will cause our product to regenerate the summary, thus training our product to improve. The system will leverage fine-tuned LLMs with LoRA optimization for efficiency.

To build the system, we will integrate PDF parsing tools (TBD by trial and error, scipdf_parser) and APIs (TBD: ArXIV, ..) to enhance citation analysis. The backend will process text and generate summaries, hosted on Github Pages or Google Cloud will serve as the deployment platform. The main technical challenge will be ensuring quality summarizations with good flow, optimizing inference for speed, and handling diverse research paper formats, which we will address through fine-tuning model pruning, and real-world data set curation.

5 Evaluation Plan

To evaluate the success of our project we will apply a hybrid of NLP metrics and user engagement data. To address our model performance we use commonly used metrics in NLP such as ROUGE, BLEU, and METEOR scores to identify the increase in performance after applying LORA to our base model. We justify the use of ROUGE to measure our summary outputs, as ROUGE is designed specifically for that reason Auriemma Citarella et al. [2004]. The use of BLEU and METEOR are designed to make sure our model is still outputting human-readable output after fine-tuning training Lavie and Denkowski [2009] and Reiter [2018]. We also plan to apply user feedback on generated summary outputs to help further guide future training, as well as to help justify overall model performance.

References

Alessia Auriemma Citarella, Marcello Barbella, Madalina G Ciobanu, Fabiola De Marco, Luigi Di Biasi, and Genoveffa Tortora. Rouge metric evaluation for text summarization techniques. *Available at SSRN 4753220*, 2004.

Alon Lavie and Michael J Denkowski. The meteor metric for automatic evaluation of machine translation. *Machine translation*, 23:105–115, 2009.

Ehud Reiter. A structured review of the validity of bleu. *Computational Linguistics*, 44(3):393–401, 2018.

SciSummary. Use ai to summarize scientific articles and research papers in seconds, 2025. URL https://scisummary.com/. Accessed 4 Feb. 2025.