LLMResearch: Large Language Model Research Toolkit

A project report

submitted by

Ayush Kumar(IIT2022090)

Avinash Kumar Nirmal(IIT2022152)

Ankit Prakash(IIT2022222)

under the supervision of

Dr. Naveen Saini

in partial fulfillment of
the requirements for the degree of
Bachelor of Technology (Information Technology)

to the

Department of Information Technology
Indian Institute of Information Technology, Allahabad

september 2024

Indian Institute of Information Technology, Allahabad
DEPARTMENT OF
Information Technology

DECLARATION

I hereby declare that the work reported in the project titled **"LLMResearch: Large Language Model Research Toolkit"** submitted for the partial fulfillment of M.Tech. degree at the Department of Information Technology, Indian Institute of Information Technology, Allahabad, is record of my work carried out under the supervision of *Dr. Naveen Saini.*

Ayush kumar(IIT2022090)
Avinash Kumar Nirmal(IIT2022152)
Ankit Prakash(IIT2022222)
(B.Tech. Pre-Final)
Department of Information Technology
IIIT Allahabad
September 2024

ACKNOWLEDGEMENT

As a matter of first importance, I offer my genuine thanks to my supervisor Dr. Naveen Saini, Associate Professor, IIIT Allahabad. I appreciate his support and help during the project work. Own Ideas/Acknowledgement

Ayush Kumar(IIT2022090)

Avinash Kumar Nirmal(IIT2022152)

Ankit Prakash(IIT2022222)

Contents

1 Introduction **2**

1.1 What is Literature Review 2

1.2 What is NLP 2

1.3 What are Large Language Models 2

1.4 What is the problem Statement 3

1.5 What is the Existing solution 3

1.6 What is the Proposed Solution 3

2 Literature review **4**

3 Report on the present investigation **6**

3.1 Research Paper Parsing using Science Parser 9

1 Introduction

1.1 What is Literature Review

A literature review is a comprehensive summary and analysis of existing research on a particular topic. It involves collecting and critically evaluating previously published work, identifying gaps in knowledge, and establishing the context for the research question or problem being investigated. The purpose of a literature review is to provide readers with a clear understanding of the current state of research in a field, while also showing how a new study will contribute to or build upon existing knowledge.

1.2 What is NLP

In the context of a research paper, Natural Language Processing (NLP) is a field within artificial intelligence (AI) that focuses on enabling computers to process, understand, and generate human language. It draws from linguistics, computer science, and machine learning to analyze and manipulate natural language data, whether written or spoken.

Relevance to Research: NLP has become essential for automating the processing and analysis of large volumes of text in fields like Literature Review. As the field evolves, techniques like Retrieval-Augmented Generation (RAG) and the use of large language models (LLMs) are at the forefront, addressing challenges such as hallucinations, knowledge updates, and domain-specific expertise, as discussed in many recent research works.

1.3 What are Large Language Models

A large language model (LLM) is an AI system designed to recognize and generate text, among other functions. These models are trained on vast amounts of data, which is why they are referred to as "large." LLMs utilize machine learning, specifically through a type of neural network known as a transformer model.

Put simply, an LLM is a software program that has been trained with extensive examples to understand and interpret human language or other complex data types.

Many LLMs are trained on vast amounts of text collected from the Internet, often amounting to thousands or millions of gigabytes. However, the quality of the data used significantly affects how well LLMs can understand and generate natural language, so developers may refine the training dataset to improve performance.

1.4 What is the problem Statement

It is difficult and time-consuming for researchers to stay updated with new publications, even with existing tools that provide daily paper feeds.

Existing tools typically generate explanations for a single research paper, but they do not capture the relationships between multiple papers.

1.5 What is the Existing solution

Past work has been able to generate single citation sentences, often in the context of one paper citing another. However, this approach does not address the need for more extensive reviews that include transitions and connections between papers.

1.6 What is the Proposed Solution

The researchers propose a feature-based approach using Large Language Models (LLMs) to generate richer and more comprehensive citation texts. Their system aims to generate not just individual citation sentences but also transition sentences, creating a coherent "story" that reflects the relationships between multiple papers.

2 Literature review

LitLLM: A Toolkit for Scientific Literature Review:The paper introduces LitLLM, a system designed to generate literature reviews using large language models (LLMs). LitLLM leverages a retrieval-augmented generation (RAG) approach with a re-ranking strategy to produce literature reviews with proper attribution. It enables researchers to quickly search and generate relevant related work based on a research idea, proposal, or abstract. The system is built on a modular pipeline, making it adaptable to future LLMs and applicable to other domains like news by modifying the retrieval source. The authors emphasize that, while LitLLM shows potential as a research assistant, it is crucial to disclose its usage in research papers and remain cautious about possible hallucinations from LLMs. Future work aims to improve the system by incorporating full academic papers instead of abstracts, using advanced LLMs and PDF parsing tools. The ultimate goal is to develop more intelligent and interactive research assistants that can better support researchers in generating accurate and comprehensive literature reviews.

The paper "Explaining Relationships Among Research Papers" by Xiangci Li and Jessica Ouyang presents a novel approach to generating related work sections and literature reviews by leveraging Large Language Models (LLMs). This work aims to address key challenges faced in automatic literature review generation, particularly the need to identify the relationships among multiple cited papers. Traditional citation generation models primarily focus on single citation sentences in isolation, failing to capture the narrative transitions that connect different research papers cohesively.

The authors introduce a feature-based approach, where LLMs like GPT-4 are used to generate citation texts enriched with transition sentences that provide coherence and structure to multi-document summaries. They propose human-interpretable features, such as the relationship between cited papers, citation intent, and writing style, which guide the LLMs in generating a connected story rather than isolated summaries. Expert evaluations indicate that readers prefer a more integrative writing style that abstracts high-level information and uses transitional sentences to present a coherent

overall narrative.

By utilizing advanced LLMs and focusing on multi-citation generation, the paper demonstrates a significant improvement in the quality of automatically generated literature reviews. This approach makes it easier for researchers to understand the interconnections between papers, improving the overall experience of reading and curating scientific literature, while also paving the way for further advancements in automatic citation generation.

The paper "The Use of Artificial Intelligence in Writing Scientific Review Articles" by Melissa A. Kacena, Lilian I. Plotkin, and Jill C. Fehrenbacher explores the utility of AI, specifically ChatGPT, in drafting scientific review articles. The study evaluates three approaches: human-only, AI-only, and AI-assisted writing to determine AIâs effectiveness in reducing the time and effort required for manuscript creation. The authors tested these approaches on three musculoskeletal research topics and found that while AI reduced writing time, it introduced significant inaccuracies, particularly in references, with up to 70 percentage being incorrect in the AI-only method. Additionally, AI-assisted writing resulted in higher plagiarism similarity scores compared to the human-only approach. Despite these issues, the study suggests that AI, when used with careful human oversight, can assist in the scientific writing process, though it is not yet capable of independently producing accurate, publishable work.

3 Report on the present investigation

LitLLM: A Toolkit for Scientific Literature Review: We have closely followed Shubham Agarwal, Issam H. Laradji, Laurent Charlin, Christopher Pal1 work, given his significant contributions to this field, and have thoroughly studied the limitations and future directions he outlined, allowing us to analyze potential next steps for further advancements.

By reading and Analysing this paper we understand about problem statement and the solution to that problem and future scope . LLM-based Re-ranking: The system uses an LLM-powered re-ranking approach to identify the most relevant research papers related to a provided abstract. This ensures that the retrieved papers align closely with the topic and provide useful contextual knowledge.

LLM-based Generative Model: Based on the re-ranked papers and the user-provided summary, LitLLM employs an LLM generative model to produce a literature review. This generation can be optionally controlled by a sentence plan to ensure a logical structure and coherence.

Contextual Knowledge for Citation: The tool enables accurate citation by retrieving existing papers and incorporating relevant contextual knowledge into the LLM-based generation process, thus maintaining factual accuracy and reducing hallucinations in the generated content.

Modular Pipeline System: LitLLM is designed as an interactive, modular pipeline, allowing scientists to write a literature review or related work section starting from an abstract. The system helps users navigate through the vast amount of research papers by offering a structured approach.

Retrieval-Augmented Generation (RAG): The system uses RAG techniques to ensure that the generated literature review is factually grounded. By incorporating multiple search strategies, the system minimizes the risk of introducing errors or hallucinations.

Sentence-Based Planning: The inclusion of sentence-based planning in the generation process offers controllable output, making it easier for users to customize the structure and flow of the generated literature review according to their needs.

Explaining Relationships Among Research Papers : The learnings from this paper is as follows:–

Challenge of Staying Updated: Keeping up with the rapid pace of research publications, including pre-prints, is a time-consuming task for researchers. Existing tools like daily research feeds help, but they still require manual curation, reading, and digestion of numerous papers. Therefore, there is a pressing need for concise, automatically generated literature reviews that can summarize these new papers and be customized for individual researchers.

Lack of Suitable Datasets: There is currently no dataset specifically designed for automatically generating customized literature reviews or summaries of daily research feeds. While survey articles exist, they are often too long and not tailored to a specific reader. The authors suggest using related work sections of scientific papers as a proxy for these concise, customized summaries because they are short, customized, and readily available.

Existing Citation Generation Approaches: Previous approaches to citation generation, both extractive and abstractive, have typically focused on generating individual citation sentences in isolation. However, these methods fail to capture the relationships between cited papers, which are crucial for crafting a coherent literature review or related work section.

Need for Multi-Document Summarization: Effective literature reviews or feed summaries should be multi-document summaries that include not only citation sentences but also expository and transition sentences to weave together the cited papers into a coherent narrative. Current methods, including neural sequence-to-sequence models, struggle to generate these multi-document summaries due to length constraints and the inability to use supporting features like citation intents and topic information.

LLM Limitations and Prompting-Based Methods: While large language models (LLMs) like GPT-4 show promise in generating richer citation texts, they still struggle with factual accuracy and producing coherent, on-topic literature reviews. The paper highlights the limitations of current LLMs, which can hallucinate citations or produce generic expository sentences.

Feature-Based, LLM-Prompting Approach: The authors propose a feature-based approach to guide LLMs in generating more accurate and structured literature re-

views. By using features that capture relationships between cited papers and the citing paper, they develop a new prompt that can generate multiple citations and transition sentences in one pass, improving the coherence of the generated paragraphs.

Planning-Based Generation: In this work, the authors experiment with using a high-level plan to guide the generation of citation paragraphs. By providing a few sentences describing the relationships among cited papers, they show that the overall organization of the generated literature review improves.

Expert Evaluation: An expert evaluation revealed that readers prefer literature reviews with more abstract, high-level citations and transition sentences that connect the papers into a coherent story. The study found a strong correlation between human preference and an integrative writing style, suggesting that this approach leads to higher-quality, more readable literature reviews.

The Use of Artificial Intelligence in Writing Scientific Review Articles: Potential of AI in Scientific Writing: The study emphasizes how AI, particularly LLMs like ChatGPT, can potentially save time in scientific writing by automating parts of the literature review, outlining, and drafting process. AI can analyze sources much faster than humans and produce readable, coherent text.

Challenges with AI in Writing: Despite its potential, AI also faces several challenges:

Artificial Hallucinations: AI can generate falsified or inaccurate information, confidently presenting it as fact. This presents a risk when relying on AI to generate scientific content. Bias and Inaccuracy: AIâs outputs can be biased or inaccurate due to the data it was trained on or the lack of critical evaluation of sources. Citation Issues: AI models, like ChatGPT, have been known to generate fictitious citations, leading to credibility issues in scientific writing. AI in Scientific Writing: The paper discusses how AI is increasingly being used in scholarly writing but notes that it should be used cautiously. While AI can help with grammar, vocabulary, and writing style, misuse can lead to serious consequences, such as generating false information or violating copyright laws.

Three Writing Strategies:

Human-Only (Traditional Method): This process involves researchers conducting the literature review, outlining, and writing the review article themselves. It takes

the most time but is expected to require the fewest revisions. AI-Only: AI handles the entire process, from literature review to drafting. This is the fastest method but expected to produce the most errors and require the most changes due to the risks of hallucinations and inaccuracies. AI-Assisted: Humans conduct the literature review and outline, and AI is used to draft the article. This approach is expected to strike a balance, saving time while requiring fewer revisions than the AI-only method. AI's Role in Enhancing Research Efficiency: While AI cannot replace human researchers, it can enhance efficiency by automating tedious tasks like sifting through literature and drafting. Particularly for non-native English speakers, AI can help with grammar, vocabulary, and writing structure.

Human Oversight Remains Crucial: The study underlines the importance of human involvement in the process. Even with AIâs capabilities, researchers need to manually validate the information, carefully select references, and ensure the accuracy of the review. This helps mitigate AI's limitations and improve the quality of the final output.

Improved Process through Human-AI Collaboration: The AI-assisted approach highlights the potential for AI-human collaboration in scientific writing. By combining human expertise in research with AIâs efficiency in generating text, this hybrid method can save time while maintaining quality.

Ethical and Practical Considerations: The study also touches on the ethical implications of AI in academic writing, urging scrutiny and caution. Concerns over AI-generated content include intellectual dishonesty, potential copyright violations, and the risk of perpetuating low-quality publications filled with errors or false information.

3.1 Research Paper Parsing using Science Parser

Parsing research papers involves extracting structured data from academic documents to analyze and utilize their content. In our implementation, we utilized the Science Paper tool to parse two research papers, automating the extraction process using Docker. Docker provides a controlled environment where the parsing tool can run efficiently, ensuring reproducibility and isolation from system-specific dependencies.

Once the papers are parsed, the extracted data is output in a structured .json file,

which includes critical components like metadata, section headings, references, and citations. This .json format serves as the foundation for further analysis such as in generating related works. By analyzing the structured data from the parsed research papers, the extracted references and citations can be compared and synthesized to construct a coherent related works section.

[hidelinks]hyperref

References

[1] Melissa A. Kacena, Lilian I. Plotkin, and Jill C. Fehrenbacher, *The Use of Artificial Intelligence in Writing Scientific Review Articles*, Current Osteoporosis Reports, 2024.

[2] Xiangci Li and Jessica Ouyang, *Explaining Relationships Among Research Papers*, arXiv, 2024.

[3] Shubham Agarwal, Issam H. Laradji, Laurent Charlin and Christopher Pal, *LitLLM: A Toolkit for Scientific Literature Review*, arXiv, 2024.