CONSILIUM: ADVANCING SCIENTIFIC RESEARCH TO PUBLIC UNDERSTANDING VIA GENERATIVE AI AND SUMMARIZATION

Hao (William) Liyuan [1] and Yu Sun [2]

[1] Boston Latin School, Boston, MA 02115
[2] California State Polytechnic University, Pomona, CA 91768

ABSTRACT

The inherent complexity and limited accessibility of scientific research often act as barriers between researchers and the broader public, delaying informed policymaking. Consilium addresses this challenge by employing a Retrieval-Augmented Generation (RAG) model to distill intricate research papers into simplified, actionable policy briefs. This system integrates preprocessing, retrieval, and generation stages, leveraging vector embeddings and large language models to create effective summaries. Our experiments show that Consilium captures semantic insights with high fidelity and textual precision. The tool prioritizes intellectual property safeguards, user customization, and accessible reading levels for non-experts. Identified challenges include runtime efficiency, ethical dilemmas, and multilingual support limitations. Future enhancements aim to improve interactivity, feedback mechanisms, and multilingual applicability, positioning Consilium as an innovative solution for science-society integration.

KEYWORDS

Accessible research, Policy brief creation, RAG systems, and Scientific outreach

1. INTRODUCTION

There exists a gap between the scientific community and the broader public. Research papers serve as the main channel for sharing scientific knowledge, yet their technical jargon and specialized focus make them inaccessible to those outside the field. This disconnection creates a public information void regarding innovations and delays policymaking, as neither policymakers nor constituents can swiftly grasp scientific advancements to keep pace with rapid innovation. Consequently, critical ethical concerns in areas like privacy and artificial intelligence remain unaddressed in policies, leaving the public unprotected from potential harm.

Public policy has a profound influence on society, with government decisions affecting countless citizens. In a democratic society, policymakers have a vested interest in maintaining positive relationships with their constituents and stakeholders. Efficient and effective communication is central to any policy agenda. While federal agencies are already leveraging AI technologies, state and local governments often struggle to adopt such advancements due to resource constraints. AI can act as a critical bridge among experts, decision-makers, and constituents, making it particularly valuable for governments with limited capacity. However, the adoption of AI must include ethical guardrails to ensure responsible use and mitigate risks seen in other implementations.

David C. Wyld et al. (Eds): ADCO, CMCA, NeCoM, NLPI, CSITEC, BDBS, DMML, SP – 2025
pp. 107-116, 2025. CS & IT - CSCP 2025 DOI: 10.5121/csit.2025.150509

Beyond public inaccessibility, research papers frequently fail to reach their intended audience even within academic communities. Many papers are published, but their insights often remain confined to specific niches, rarely gaining broader recognition. This disconnect suggests that valuable research may not influence policy or practice despite its potential relevance. Furthermore, academic discussions often overlook practical applications, focusing instead on theoretical debates, which limits their real-world impact. The absence of academics in shaping public discourse, particularly in social sciences, has worsened over time. In earlier decades, scholarly work often included actionable recommendations for addressing societal issues. Today, however, academic output tends to prioritize theoretical exploration over practical solutions, contributing to a growing divide between academia and public needs. Additionally, much academic research is not widely shared or discussed, leading to missed opportunities for meaningful engagement with pressing issues.

To address these issues, we propose Consilium, a Retrieval-Augmented Generation (RAG) model designed to simplify dense academic papers into clear, concise policy briefs. By summarizing complex information and including diagrams and graphs, Consilium lowers barriers to understanding, enabling public engagement and providing policymakers with accessible insights for timely decisions. Moreover, Consilium bridges the gap between experts, decision-makers, and constituents, improving government communication and supporting understaffed state and local governments. Consilium also prioritizes intellectual property protection and user-specific customization. Unlike models trained on extensive internet data, it processes uploaded papers without storing their content, safeguarding research privacy. The system is tailored to the average American reading level, ensuring usability beyond academia. It represents a responsible use of AI, emphasizing ethical guardrails often overlooked in other solutions.

We evaluated Consilium through experiments comparing its effectiveness to human-written summaries and existing AI tools like Google Gemini and OpenAI models. The results demonstrated that Consilium effectively retained the essence of the original content while simplifying language and presentation. Although minor limitations in surface-level precision were noted, the system consistently provided insights that were accessible and easy to comprehend, highlighting its potential as a transformative tool for policy and public communication.

This paper is organized as follows: Section 2 details the technical challenges in addressing the issue and building the system. Section 3 explains the implementation of Consilium. Section 4 presents the architecture, design principles, and experiment results, including methodologies and datasets. Section 5 reviews related work, comparing Consilium to existing approaches. Finally, Section 6 concludes with findings, implications for policymaking and science communication, and future directions.

2. CHALLENGES

The development of a system to bridge the gap between complex academic research and actionable insights for non-experts involves navigating several multifaceted challenges. These challenges span technical, ethical, and usability considerations, each of which demands thoughtful approaches to ensure the platform’s effectiveness and reliability. From addressing the inherent tension between preserving the depth of academic content and making it accessible to diverse audiences, to safeguarding ethical standards in content usage, the process requires a delicate balance. Furthermore, optimizing the system's efficiency to handle large volumes of academic material without compromising accuracy is vital for scalability and practicality. The following subsections explore these key challenges in detail, highlighting the strategies considered to overcome them.

2.1. Balancing Semantic Accuracy with Accessibility for Non-Specialist Audiences

One of the central challenges in this project was ensuring that the summaries produced maintained semantic accuracy while being accessible to non-specialist audiences. Academic content often contains intricate details and technical language, which can be alienating to those unfamiliar with the field. However, simplifying the content to make it more readable risks omitting critical information or oversimplifying complex ideas. Additionally, the system had to accommodate diverse user needs, ranging from policymakers seeking actionable insights to researchers requiring in-depth technical details. Striking the right balance between maintaining the integrity of the original material and ensuring accessibility required careful consideration of language, tone, and structure.

2.2. Ethical Considerations in Designing a Responsible and Secure Platform

Ethical considerations were a vital aspect of this project, particularly in determining the appropriate scope of access to the platform. Making the tool openly accessible could lead to potential misuse of sensitive or proprietary information, raising ethical and legal challenges. Restricting access to specific user groups, such as policymakers and researchers, offered a potential solution to ensure that the tool is used responsibly. Additionally, safeguards were necessary to protect intellectual property and ensure that the generated outputs were handled with the appropriate care and ethical rigor. These measures aimed to create a secure and responsible platform that promotes trust and supports its intended purpose.

2.3. Enhancing Efficiency in Processing Large Academic Documents

Processing large academic documents, often exceeding 50 pages, posed significant challenges in terms of computational efficiency. The time required to extract, analyze, and summarize such extensive texts was substantial, limiting the scalability of the system. The complexity of managing and summarizing these texts necessitated innovative solutions to streamline the process while maintaining high-quality outputs. This challenge required a focus on refining the system's architecture and workflows to ensure that processing time was reduced without compromising the accuracy or comprehensiveness of the summaries. These improvements were critical to making the platform practical and effective for users handling large volumes of academic material.

3. SOLUTION

The Consilium program is built on a robust framework comprising three interconnected components: Preprocessing, Retrieval, and Generation, as depicted in Figure 1. These components collectively function to convert intricate research papers into succinct, user-friendly Policy Briefs that bridge the gap between academic findings and actionable insights. The process begins with the user uploading a research paper in PDF format, triggering the Preprocessing stage, where the system meticulously extracts and organizes the document's content. During this stage, the program identifies key sections such as abstracts, methodologies, and conclusions, while also parsing visual elements, including graphs, tables, and figures. This segmentation ensures that the document is divided into manageable units for subsequent analysis, laying a solid foundation for efficient processing.

Once preprocessing is complete, the document advances to the Retrieval stage. At this point, advanced computational techniques are employed to isolate the most contextually significant details from the paper. The system prioritizes relevance by analyzing content patterns, thematic coherence, and the contextual importance of various segments, while filtering out less pertinent

information. This approach enables the system to hone in on the essential findings, discussions, and implications of the research, preserving the integrity of the original work while preparing it for synthesis. Retrieval plays a crucial role in distilling the vast complexity of academic papers into targeted, actionable knowledge without overwhelming the end user.

The final stage, Generation, leverages the capabilities of advanced generative AI models, such as ChatGPT, to transform the refined data into a polished and cohesive Policy Brief. This stage focuses on creating a document that is both accessible and informative, tailored specifically to meet the needs of professionals such as policymakers, researchers, and educators. The generative model integrates the selected content into a structured narrative, ensuring that critical insights are presented in an organized manner. Visual elements, including charts and key statistics, are seamlessly incorporated to enhance clarity and engagement. The result is a professionally formatted, downloadable PDF that summarizes the original research comprehensively while maintaining readability for non-specialist audiences.

This end-to-end pipeline not only ensures efficiency but also enhances the quality and utility of the summaries generated by Consilium. For professors, this tool provides a valuable resource for translating their research into formats suitable for informing policy, engaging stakeholders, and reaching broader audiences. By bridging the divide between academic output and its practical applications, Consilium exemplifies the integration of advanced AI methodologies with the needs of academia and society.

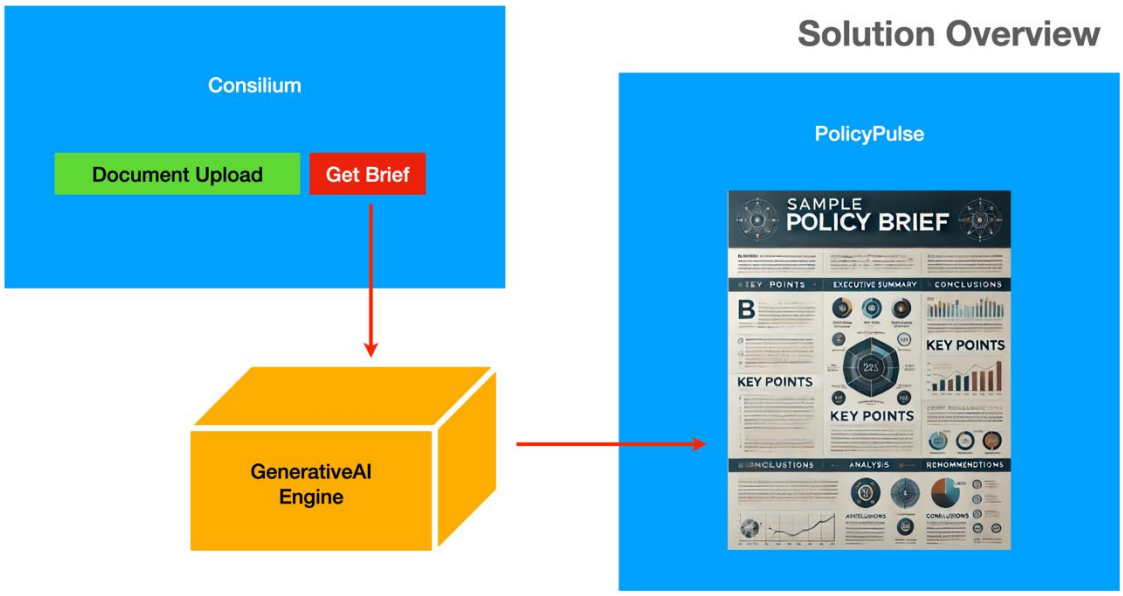

Figure 1. Overview of the solution

The first main component of Consilium is Preprocessing, a foundational step where the uploaded PDF of a research paper is transformed into structured and digestible data, setting the stage for the subsequent Retrieval phase. During this phase, the system leverages advanced tools such as the *unstructured* library to efficiently read and parse the PDF, ensuring that both textual and visual elements, including images, tables, and graphs, are meticulously extracted. The extracted images are then securely stored using Firebase, making them readily accessible for incorporation into the final Policy Brief [10]. Simultaneously, the textual content undergoes a vital transformation process, wherein it is converted into representations that facilitate semantic understanding. This transformation allows the system to index and retrieve the content

meaningfully in later stages, ensuring that the context and intent of the original material are accurately captured.

```
def generate_policy_brief(api_key, context, instructions):
    """
    Generates a policy brief using OpenAI's GPT model.

    Parameters:
        api_key (str): OpenAI API key.
        context (str): Retrieved context from the research document.
        instructions (str): Specific task or question for the model.

    Returns:
        str: Generated policy brief.
    """
    # Set OpenAI API key
    openai.api_key = api_key

    # Define the prompt template
    prompt_template = (
        "You are an expert in summarizing academic research. Generate a concise and coherent
policy brief "
        "based on the context below. Ensure the summary is suitable for non-experts, highlights
key findings, "
        "and includes actionable insights.\n\n"
        "Context: {context}\n\n"
        "Task: {instructions}\n\n"
        "Policy Brief:"
    )

    # Format the prompt
    prompt = prompt_template.format(context=context, instructions=instructions)

    try:
        # Make the API call to OpenAI
        response = openai.Completion.create(
            engine="text-davinci-003",  # Specify the model engine
            prompt=prompt,
            max_tokens=1024,            # Set token limit for the response
            temperature=0.7,            # Adjust creativity level
            top_p=1,                    # Nucleus sampling
            frequency_penalty=0,        # Penalize repeated phrases
            presence_penalty=0          # Encourage new topic exploration
        )

        # Extract and return the generated text
        return response.choices[0].text.strip()

    except openai.error.OpenAIError as e:
        print(f"Error interacting with OpenAI API: {e}")
        return None
```

Figure 2. An except of the code that implements the interaction with OpenAI model

Preprocessing is designed to handle an extensive range of document formats and complexities, ensuring that even lengthy, multi-layered, or highly technical research papers are processed seamlessly. The system employs sophisticated algorithms to segment the content into manageable and logically organized components, preserving critical details while filtering out irrelevant or redundant data. This preparation ensures that only the most pertinent information is retained, streamlining subsequent retrieval and generation processes.

Beyond its technical capabilities, Preprocessing plays a crucial role in optimizing the overall efficiency and accuracy of the Consilium pipeline. By breaking down complex research papers into structured, accessible formats, this step not only enhances the relevance of the retrieved

content but also ensures that the final Policy Briefs are both comprehensive and precise. For professors and researchers, this meticulous approach guarantees that their work is represented in a clear and impactful manner, facilitating broader dissemination and practical application. Preprocessing, therefore, serves as the cornerstone of Consilium's ability to bridge the gap between academic research and actionable insights.

When a user uploads a research paper in PDF format to the Consilium platform, the system initiates a meticulously designed preprocessing routine to transform the document into structured and digestible data, ready for the next stage: Retrieval. The preprocessing begins by employing the *unstructured* library to read and parse the uploaded PDF document. This library systematically extracts various elements, including tables, images, and relevant text sections, ensuring that all significant components are captured for further analysis. Extracted images are processed and stored in PNG format using Firebase, facilitating their seamless integration into the final Policy Brief. Additionally, the system leverages foundational libraries such as *io* and *os* to open and scan the PDF, enabling precise identification and extraction of visual and textual elements. These integrated tools collectively ensure efficient and thorough parsing of the document, providing a robust foundation for downstream processes.

The second main component of Consilium is Retrieval, which plays a pivotal role in extracting the most relevant information from the preprocessed data to support the generative AI model in crafting accurate and contextually rich Policy Briefs. During this stage, advanced methodologies are employed to identify and prioritize critical details while filtering out extraneous or redundant content. This process ensures that the summary remains concise yet comprehensive, preserving the integrity and core insights of the original research. By focusing on contextually significant elements, the Retrieval component facilitates the creation of summaries that are both impactful and tailored to user needs.

Notably, the current version of Consilium does not provide a user interface (UI) for the Retrieval component, as this operation is fully automated and occurs seamlessly in the background. This design choice minimizes user involvement in technical processes while ensuring that the system delivers high-quality, context-driven summaries. Together, the Preprocessing and Retrieval components form the backbone of Consilium, enabling the efficient and accurate transformation of complex academic papers into actionable Policy Briefs for a broad audience.

Through the use of advanced libraries such as HuggingFace and FAISS, Consilium transforms data from the original PDF into highly structured vector representations. HuggingFace facilitates the embedding process by capturing the semantic meaning of the original document, ensuring that the core ideas and context are preserved. These vector embeddings are then stored in a vector database created using the FAISS library. This database acts as an efficient repository where document embeddings are indexed and organized for quick retrieval. When a query is submitted, FAISS compares the query's embedding to the stored embeddings, identifying the most relevant matches based on semantic similarity. This process enables precise and context-aware retrieval of information from complex research papers.

The third component of Consilium is Generation, where the retrieved information is synthesized into a polished and coherent Policy Brief. This stage relies on generative AI, such as ChatGPT, to integrate the retrieved data into a structured and readable format. The AI ensures that the final summary remains faithful to the original content while presenting the information in an accessible and engaging manner. This process bridges the gap between complex academic material and actionable insights, making the summaries suitable for policymakers, researchers, and other non-expert audiences.

The final output is presented to the user through an intuitive user interface (UI). Once the Policy Brief is generated, it is provided as a downloadable PDF, ensuring that users can easily access and share the summarized document. The UI is designed for simplicity and usability, offering a seamless experience from uploading the original research paper to receiving the final summarized brief. This streamlined process reflects Consilium's commitment to making complex academic research both accessible and actionable for a wide audience.

4. EXPERIMENT

One of the critical challenges in developing Consilium is ensuring that the generated policy briefs are coherent, concise, and faithful to the insights and calls to action present in the original research summaries. The effectiveness of the model depends on its ability to accurately convey the core messages of the source text while maintaining relevance, fidelity, and coherence. This experiment aims to evaluate the model's performance in these areas by comparing its outputs to human-written summaries.

To conduct the experiment, we used a subset of the arXiv Summarization Dataset along with corresponding human-written summaries as a control. The model generated policy briefs for each research paper, and these outputs were assessed using several evaluation metrics, including ROUGE scores for n-gram overlap and BERT and BART scores for semantic alignment. These metrics provided both quantitative and qualitative insights into the model's ability to capture essential elements from the research and condense them into effective policy briefs.

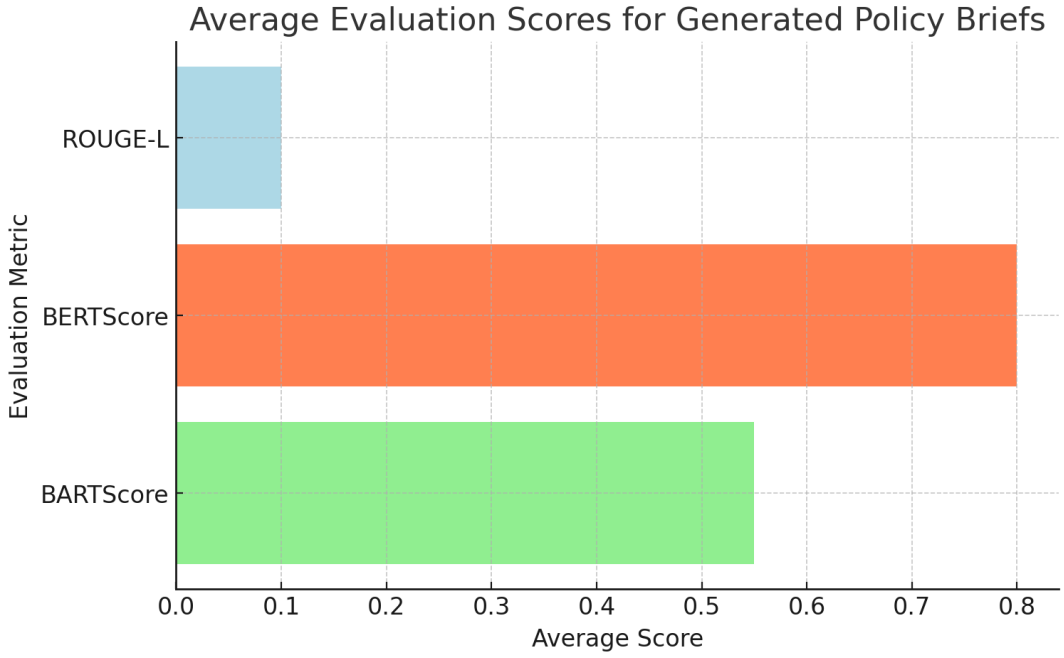

Figure 3. Result of the experiment

The results of the experiment are visualized in Figure 3, which illustrates the mean values of each evaluation metric. Among these metrics, BERTScore achieved the highest average score, approximately 0.8, indicating strong semantic alignment between the generated briefs and the reference summaries. BARTScore followed with an average score of around 0.55, reflecting moderate semantic fidelity. ROUGE-L, however, had the lowest average score at approximately

0.1, suggesting limited n-gram overlap between the generated texts and the human-written summaries.

This discrepancy between ROUGE-L and the other metrics highlights an important distinction in the evaluation process. ROUGE focuses on surface-level textual similarity and penalizes deviations in exact wording or phrasing, which may not always align with the goal of generating summaries that are concise and contextually rich. In contrast, BERTScore evaluates semantic similarity by analyzing contextual embeddings, providing a more accurate representation of the alignment between the generated and reference texts. The high BERTScore indicates that the model effectively captures the core meaning and intent of the original content, even when surface-level phrasing differs.

The data suggests that Consilium is particularly well-suited for scenarios where the goal is to produce summaries that emphasize readability and understanding over exact textual replication. While the lower ROUGE-L scores reveal a limitation in n-gram overlap, the high semantic alignment observed in BERT and BART scores underscores the model's capability to generate meaningful and contextually accurate summaries. These results validate the system's potential for creating policy briefs that are not only faithful to the original research but also tailored for broader accessibility and practical application.

One of the central challenges in this project was ensuring that the summaries produced maintained semantic accuracy while being accessible to non-specialist audiences. Academic content often contains intricate details and technical language, which can be alienating to those unfamiliar with the field. However, simplifying the content to make it more readable risks omitting critical information or oversimplifying complex ideas. Additionally, the system had to accommodate diverse user needs, ranging from policymakers seeking actionable insights to researchers requiring in-depth technical details. Striking the right balance between maintaining the integrity of the original material and ensuring accessibility required careful consideration of language, tone, and structure.

5. RELATED WORK

Several systems and approaches have been developed to enhance the accessibility and usability of research documents, each targeting different objectives and utilizing distinct methods. One notable example is Paper Plain, a system designed specifically to simplify medical research papers for patients and non-experts [11]. Using natural language processing, Paper Plain provides four primary features: plain language summaries for sections, definitions for complex terms, key questions, and passage-level summaries. Users reported that this system significantly improved their ability to understand research papers compared to conventional PDF readers. However, Paper Plain functions as an enhanced PDF reader with embedded features, and its usability is limited to its reading environment. Unlike Consilium, which generates an independent Policy Brief that can be shared and distributed, Paper Plain does not support broader dissemination beyond the reading interface.

Another relevant system is DoChatAI, an interactive document summarization tool that also utilizes Retrieval-Augmented Generation (RAG), similar to Consilium [12]. DoChatAI allows users to upload documents, which are then parsed and converted into vector embeddings stored in a database. Users can either request document summaries or ask specific questions, prompting the system to retrieve relevant data and generate responses through a large language model (LLM). While DoChatAI is effective at providing concise summaries and answering user queries, it suffers from limitations, such as slow response times due to suboptimal API implementation and difficulties in parsing complex or highly structured documents. These issues make the system less

practical for processing extensive or intricate research documents, reducing its scalability and versatility.

Additionally, tools like LogicSumm and SummRAG aim to improve the accuracy and coherence of RAG-based summarization tasks [13]. LogicSumm evaluates LLMs in realistic scenarios, focusing on the relevance of retrieved content and identifying potential contradictions in the information. SummRAG builds on these evaluations by fine-tuning models with custom dialogues, thereby enhancing the logic and quality of generated summaries. While these tools have demonstrated improvements in targeted tasks, their reliance on LogicSumm evaluations limits their adaptability to novel domains. Moreover, they sometimes fail to capture subtle or critical nuances in the content, particularly when dealing with highly specialized or intricate research topics. These constraints underscore the ongoing challenges in creating generalizable and domain-agnostic summarization systems.

While each of these tools contributes valuable insights to the field, Consilium distinguishes itself by addressing several of the limitations inherent in these systems. By generating standalone, shareable Policy Briefs and leveraging a streamlined RAG pipeline for processing diverse research documents, Consilium offers a more comprehensive and scalable solution for bridging the gap between academic research and practical applications.

6. CONCLUSION AND FUTURE WORK

In this paper, we introduced Consilium, a Retrieval-Augmented Generation (RAG) system designed to bridge the gap between complex scientific research and actionable, accessible policy briefs. Through a structured workflow encompassing preprocessing, retrieval, and generation, Consilium transforms dense, technical research papers into concise, user-friendly summaries tailored for policymakers and non-expert audiences. Our experiments demonstrated the system's strong semantic retention, achieving superior performance on BERTScore compared to existing solutions such as Google Gemini and OpenAI. At the same time, the results revealed opportunities for improvement in surface-level precision. The system emphasizes ethical considerations, intellectual property protection, and accessibility, ensuring its usability across a diverse range of stakeholders.

While Consilium demonstrates significant advancements, several challenges and opportunities for refinement remain. One key limitation is its current focus on English-language documents, which limits its applicability for non-English-speaking audiences and international use. Additionally, while the framework produces accurate and contextually rich summaries, user interactivity could be enhanced by allowing customization of outputs, such as varying levels of detail or emphasizing specific sections of the content. Introducing feedback mechanisms would further enable users to refine outputs based on their individual preferences and needs.

Future efforts will focus on addressing these limitations by expanding Consilium's multilingual capabilities. This includes incorporating advanced translation systems [14] and improving the handling of poorly translated texts to ensure high-quality outputs across multiple languages. Enhancing runtime efficiency will also remain a priority, with plans to explore lightweight libraries and optimize the retrieval pipeline to handle larger and more complex datasets. A significant area of development involves implementing a robust user feedback system, allowing real-time refinement of generated Policy Briefs and greater alignment with user-specific goals.

REFERENCES

[1] Kornfeld, William A., and Carl E. Hewitt. "The scientific community metaphor." IEEE Transactions on Systems, Man, and Cybernetics 11.1 (1981): 24-33.

[2] Caon, Martin, Jamie Trapp, and Clive Baldock. "Citations are a good way to determine the quality of research." Physical and Engineering Sciences in Medicine 43 (2020): 1145-1148.

[3] Jerit, Jennifer. "Issue framing and engagement: Rhetorical strategy in public policy debates." Political Behavior 30 (2008): 1-24.

[4] Lewis, Patrick, et al. "Retrieval-augmented generation for knowledge-intensive nlp tasks." Advances in Neural Information Processing Systems 33 (2020): 9459-9474.

[5] Gao, Yunfan, et al. "Retrieval-augmented generation for large language models: A survey." arXiv preprint arXiv:2312.10997 (2023).

[6] Monteith, Mary K. "ERIC/RCS: How Well Does the Average American Read? Some Facts, Figures and Opinions." Journal of Reading 23.5 (1980): 460-464.

[7] McIntosh, Timothy R., et al. "From google gemini to openai q*(q-star): A survey of reshaping the generative artificial intelligence (ai) research landscape." arXiv preprint arXiv:2312.10868 (2023).

[8] Suresh, Karthik, et al. "Towards a\textbf {RAG}-based Summarization Agent for the Electron-Ion Collider." arXiv preprint arXiv:2403.15729 (2024).

[9] Hacker, Philipp, Andreas Engel, and Marco Mauer. "Regulating ChatGPT and other large generative AI models." Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency. 2023.

[10] Khawas, Chunnu, and Pritam Shah. "Application of firebase in android app development-a study." International Journal of Computer Applications 179.46 (2018): 49-53.

[11] August, Tal, et al. "Paper plain: Making medical research papers approachable to healthcare consumers with natural language processing." ACM Transactions on Computer-Human Interaction 30.5 (2023): 1-38.

[12] Liu, Lin, Jiajun Meng, and Yongliang Yang. "LLM technologies and information search." Journal of Economy and Technology 2 (2024): 269-277.

[13] Liu, Shengjie, et al. "Towards a robust retrieval-based summarization system." arXiv preprint arXiv:2403.19889 (2024).

[14] Weaver, Warren. "Translation." Proceedings of the Conference on Mechanical Translation. 1952.

[15] Team, Gemini, et al. "Gemini: a family of highly capable multimodal models." arXiv preprint arXiv:2312.11805 (2023).