WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Unified Framework for Interactive and Verifiable Cross-Lingual Document Summarization

Satti Durga Surya Sandeep Reddy[1], Tabassum Akthar[2], Challa Prasad[3], Jagarlamudi Krishna Thrishagna[4], Ravipalli Naga Bhargav[5]

[1]Dept. of CSE(AI&ML), GMR Institute Of Technology Rajam, India 22341A4251@gmrit.edu.in
[2]Dept. of CSE(AI&ML), GMR Institute Of Technology, Rajam, India 21341A4250@gmrit.edu.n
[3]Dept. of CSE(AI&ML), GMR Institute Of Technology, Rajam, India 2341A4213@gmrit.edu.in
[4]Dept. of CSE(AI&ML), GMR Institute Of Technology, Rajam, India23345A4205@gmrit.edu.in
[5]Dept. of CSE(AI&ML), GMR Institute Of Technology. Rajam, India22341A4247@gmrit.edu.in

ABSTRACT –

Cross-lingual document understanding is a major challenge. Current workflows usually involve several separate tools for translation and summarization, but there is little insight into the accuracy of the summaries produced. This fragmented process leads to inefficiency and a significant lack of trust, limiting the practical use of AI-driven document processing. In this paper, we present a web-based framework that combines multilingual translation, abstractive summarization, and interactive verification into a single system. We use NLLB-200 for high-quality translation in over 200 languages and mT5 for summarization. Our method produces concise, human-like outputs. To ensure reliability, we provide a click-to-verify feature that links each summary sentence to its exact source in the translated document using MiniLM embeddings. We also improve accessibility with a text-to-speech "Listen" mode that allows for inclusive and multimodal interaction. The experimental evaluation and system demonstration show that our framework can increase efficiency, build user trust, and enhance accessibility. This offers a practical solution for researchers, professionals, and multilingual communities.

Keywords: Cross-lingual Document Understanding, Multilingual Translation, Abstractive Summarization, Interactive Verification, Semantic Embeddings, Human–AI Collaboration, Text-to-Speech Accessibility, Transformer Models.

I. Introduction

Breakthroughs in natural language processing NLP and machine translation have made information available in several languages. Current document understanding workflows end though, are still siloed, with most involving distinct translation and summarization tools.

The workflows typically depend on generic models that value speed and surface-level coherence, creating outputs that cannot be verified and are hard to trust. The lack of transparency leaves an important trust gap, especially in the areas of accuracy and dependability matter. Given the diversity of the clients, there's an emerging need to Intelligent systems that incorporate translation and Summarization with verifiability mechanisms. Utilizing the latest transformer-based architectures, these systems can reduce cognitive load and save analysis time, and facilitate a new level of Human– AI collaboration in which outputs are both Contextually correct and traceable to origins.

In this paper, we introduce a comprehensive web-based system which combines multilingual translation, abstractive summarization, and Interactive Verification in a single application. Our system uses NLLB-200 for fastReliable translation in more than 200 languages including mT5 for compact abstractive summarization. To build Trust, We introduce a click-to-verify functionality. driven by the MiniLM embeddings, connecting each Summary sentence with the corresponding exact source passage. We also improve accessibility with a text-to-speech "Listen" mode, offering barrier-free, multi-modal interaction.

This paper has several major contributions:

- Integrated Framework: one-stop platform with Translation, Summarization, and verification.
- Verifiable Summarization – A semantic Embedding-based system that connects summaries to their origin.

Our prototype exhibits lower cognitive load, quicker analysis, and enhanced user trust, thereby presenting a viable solution for multilingual document intelligence.

II. Literature survey

This collection of research highlights the most recent advancements in document intelligence, multilingual natural language processing, and cross-lingual summarization. Overall, they capture how researchers are working toward more accurate, verifiable, and resource-effective systems.

Presented MATSFT, an abstractive summarization system for multilingual low-resource Indian languages. Through user query fine-tuning of the mT5 model, their system generated more targeted and relevant summaries, with performance considerably enhanced for under-represented languages. [1]

Suggested a framework for cross-lingual extreme summarization of academic papers. They made comparisons between pipeline and end-to-end approaches and showed that both methods are effective but end-to-end models tend to generate better quality domain-specific summaries. [2]

Developed an approach for automatic data fetching to build large-scale cross-lingual summarization datasets. By filtering and aligning article–video pairs, they established a dataset of more than 28,000 English–Hindi pairs, which improved model training and performance. [3]

Proposed CL-XABSA, a contrastive learning model for cross-lingual aspect-based sentiment analysis. They integrated token-level and sentiment-level learning with distillation of knowledge to achieve state-of-the-art performances on multilingual sentiment analysis benchmarks. [4]

Presented a retrieval-based in-context learning system for cross-lingual summarization in low-resource languages. Their multi-language retrieval model supplied appropriate examples to large language models so that they could perform effective few-shot and zero-shot summarization. [5]

Released CrossSum, a large cross-lingual summarization dataset with 1,500+ language pairs. Multilingual model training was scaled and performance enhanced over various languages due to their retrieval and sampling techniques. [6]

Revisited the limitations of current summarization benchmarks and introduced a new dataset with better annotation quality. Their benchmark improved the faithfulness and adequacy of cross-lingual summaries, providing a stronger foundation for evaluation. [7]

Suggested multi-target cross-lingual summarization, in which a single input can produce summaries in different languages. Their language-independent approach guaranteed semantic coherence among outputs, proving the viability of this new task.[8]

Carried out a survey of visually-rich document understanding (VRDU) with deep learning. They discussed multimodal models that integrate text, vision, and layout features and presented a taxonomy of VRDU approaches and discussed open research challenges. [9]

Applied multilingual transformer-based summarization to healthcare texts. Their model outperformed in summarization accuracy in the German medical field, demonstrating the flexibility of pre-trained language models in technical domains. [10]

Explored the problem of positional bias in long-form summarization. They discovered that summaries frequently ignore the middle parts of documents.They suggested ways to ensure a more complete and accurate representation. [11]

Introduced SimCSum, a framework that combines simplification and summarization across languages. With a shared encoder multi-task model, they improved readability and overall quality of multilingual summaries, particularly in science journalism. [12]

Proposed SumTra, a differentiable pipeline for few-shot summarization across languages. Their method combined summarization and translation phases, reducing error propagation and achieving better results in low-resource conditions. [13]

Proposed a single framework that merges multilingual and cross-lingual summarization into one model. We improved generalization and scalability to multiple languages with joint training methods. [14]

These papers focus on closing language gaps, making access to datasets easier, and building trust in summarization results.In real-world multilingual contexts, there is a clear trend toward systems that are scalable, flexible, and reliable.

III. METHODOLOGY

This chapter describes the method used to design and develop the framework called “A Unified Framework for Interactive and Verifiable Cross-Lingual Document Summarization.” The system connects machine translation, summarization, and content verification by bringing together modern transformer models in a single interactive environment. Each step in the method is organized to ensure that the framework operates efficiently, keeps context accurate, and maintains user trust.

3.1 Data Collection

The success of any summarization system relies on the quality and variety of the data it uses. Therefore, the first step is to gather and choose suitable datasets for multilingual training and evaluation. To ensure broad coverage, benchmark corpora such as CrossSum, XL-Sum, and WikiLingua were employed. Each dataset has pairs of documents and summaries in various language combinations. These datasets introduce the system to a variety of grammatical structures, sentence types, and cultural details in multiple languages. Besides these public datasets, the prototype can also handle real-world documents provided by users. Users can upload files in .pdf or .docx formats through the web interface. These documents can come from various areas, including academic research, legal contracts, government policies, or healthcare reports. Allowing user-defined input keeps the system flexible and

suitable for both professional and academic uses. The data collection process includes both structured (benchmark) and unstructured (real-world) sources, creating a solid foundation for translation and summarization.

3.2 Data Pre-Processing

Before using textual data in transformer models, pre-processing is essential for maintaining consistency and accuracy. This step involves cleaning, normalizing, and segmenting text to ensure that each model, which includes the translator, summarizer, and verifier, receives high-quality input.

3.2.1 Text Extraction

This system is designed for the effective parsing of documents using the LangChain framework. It has two major loaders: PyPDFLoader for PDF files and Docx2txtLoader for Microsoft Word documents. These loaders will extract text, keeping boundaries such as paragraphs and metadata, into one continuous raw text string. Further processing can be done without loss of structure in this format.

3.2.2 Cleaning and Normalization

The text extracted from the source normally contains a lot of unwanted symbols, irregular spacing, and non-linguistic characters that hamper the tokenization process. Pre-processing scripts clean unnecessary punctuation, normalize spaces, and handle special encodings such as bullet symbols and hyphenations. Further normalizations include lowercasing, removal of common words, and sentence boundary markings. This is to make the text comply with what is required as input text by transformer models. Finally, this cleaned text is divided into manageable chunks to prevent overflow, considering the model's token limits, particularly for translation and summarization tasks.

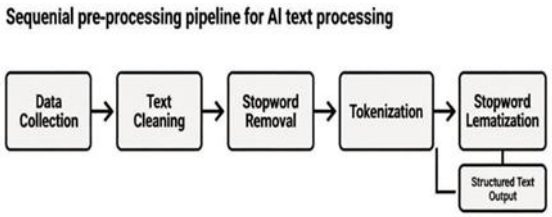


Figure 3.2 shows the pre-processing pipeline, illustrating how raw documents are turned into clean and structured text for AI processing.

3.3 Model Design

For the proposed methodology, the model design step is the most important in this regard. It fuses three transformer models for translation, summary, and verification into one. These combined models collectively constitute one intelligent and verifiable summarization system.

3.3.1 Machine Translation Module

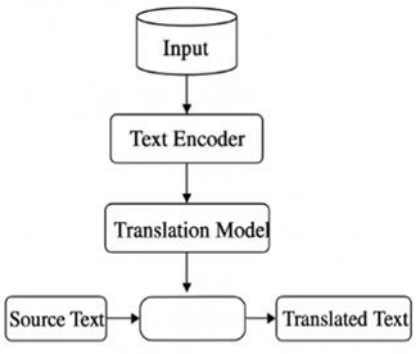


Figure 3.3 shows the structure of the translation subsystem.

This is the step in which the multilingual text is translated into a working language with the help of the machine translation phase. To perform this task, the system utilizes Llama 3 via the Groq API. Llama 3 is a large generative model that can manage complex linguistic structures while maintaining the flow of contextual meaning. This is powered by the Groq inference engine, which brings low latency-even for long documents-almost to real-time translation. Translation occurs in chunks. This means the text is divided into smaller fragments to avoid token overflow. Each chunk is translated and then combined into a single translation of the document. In this way, coherence is maintained without a sacrifice of performance. For comparison, the NLLB-200, or No Language Left Behind, model from Meta AI was also evaluated. While NLLB-200 has wide coverage in 200 languages, its inference

time was longer. Thus, the Llama 3 via Groq API was found to be the best translation model for this use case due to its nice balance of speed, scalability, and accuracy.

3.3.2 Summarization Module

When the translation is finished, the document is summarized using mT5, a Multilingual Text-to-Text Transfer Transformer model. This model sees every language task as a text-to-text problem. It allows for abstractive summarization in many languages using the same approach. Unlike extractive summarization methods that pick key sentences from the text, mT5 generates new content in its own words while maintaining the main ideas and intent of the original materials.This results in summaries that sound more natural, logical, and contextually accurate. An iterative refinement mechanism ensures comprehensive coverage. First, an abstract is generated about the first segment of translated text. As more segments are processed, the model reworks and enhances the summary to insert new context. The approach avoids a common "front-loaded bias" seen in single-pass summarization, wherein each portion of the document is given equal voice in the summary.

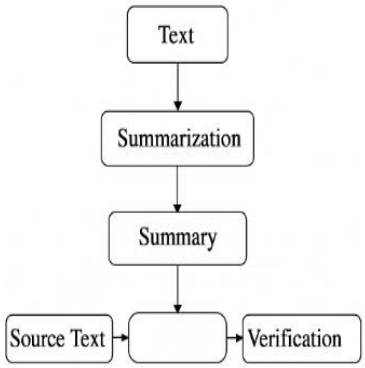


Figure 3.4 shows the summarisation process iteratively.

3.3.3 Verifiable Summarization Module

The architecture provides a verification layer that boosts users' confidence in the automatically generated content. This module verifies whether each sentence in the summary corresponds to the source document. For this, the system makes use of a paraphrase-multilingual-MiniLM-L12-v2 model: a transformer architecture that expresses sentence embeddings that capture semantic meaning. First, both the original and summarized sentences are mapped into a numerical vector representation. By means of cosine similarity, the system measures the distance of each sentence in the summary from its most similar sentence in the original document. In this way, this verification step adds traceability to the summarization process. Users understand which source sentences are related to each point, significantly improving their confidence in the reliability and accuracy of the content generated.

3.3.4 Multi-Modal Output and Accessibility

The architecture has been designed to include a multi-modal output system, considering the accessibility required for modern intelligent systems. In addition to text, the generated summary is converted to spoken audio by the Google Text-to-Speech library. This feature lets users listen to summaries, benefiting those who prefer auditory review or those with visual impairments. The final UI shows original text, translated text, and summarized text side by side to allow the user to cross-reference all three at once.

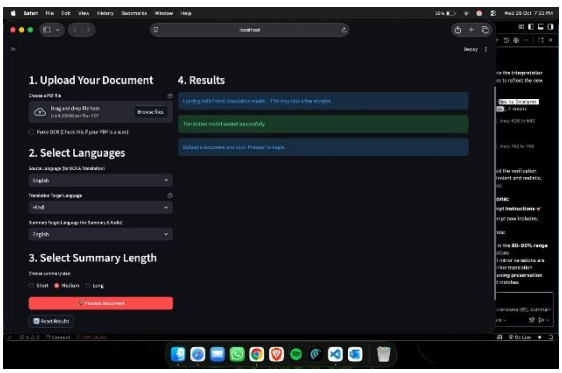


Figure 3.5 shows the interactive user interface and output visualization.

3.4 Implementation Workflow

The proposed system implements maintainability and scalability through a modular, service-based architecture. The frontend is developed with React JS to provide a responsive user experience, while the backend is maintained using FastAPI for model integration efficiently and asynchronously.An experimental interface will be created on the Streamlit framework for demonstrating a prototype. A detailed workflow for this is as follows:

- **Upload Document:** The user will upload the .pdf or .docx document through the frontend.
- **Extraction of the Text:** LangChain loaders will read and clean the text of the document
- **Translation Phase:** The clean text will be divided and translated into the required language using Llama 3 and the Groq API.
- **Summarization Phase:** The translated text will be summarized with the iterative mT5 model.
- **Verification Phase:** Semantic alignment will be checked by using MiniLM sentence embeddings and cosine similarity measures.
- **Output Generation:** Translated and summarized final results with verification marks and audio playback will be presented.

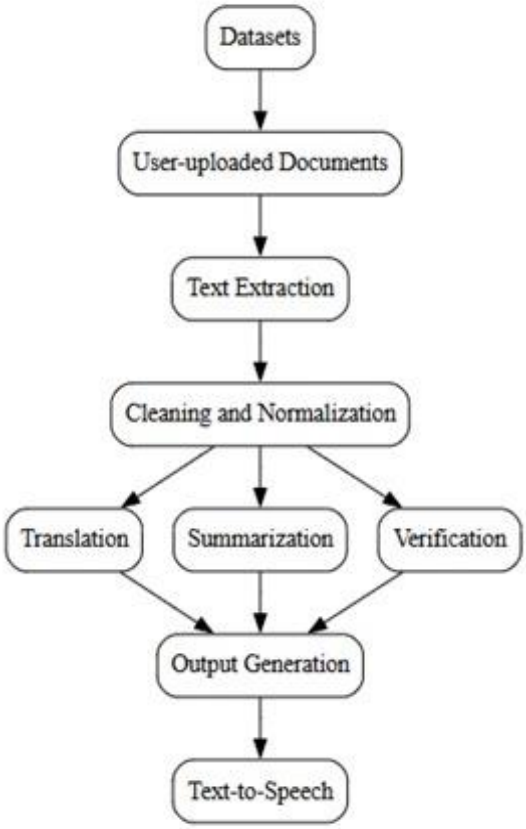


Figure 3.6 shows the detailed explanation of the implementation

Such modular work ensures real-time interaction and an efficient synchronization of the translation, summarization, and verification steps.

3.5 Evaluation Metrics

Both quantitative and qualitative methods are utilized to ensure the validity of the system's performance.

3.5.1 Quantitative Evaluation

ROUGE Score: Measured by the overlap between generated summaries and reference summaries. This is done using n-grams and sentence order.

BLEU Score: The quality of translations is scored by comparing the model's output to human reference translations.

BERT Score: It measures how similar meanings are. It does this by using cosine similarity and contextual embeddings from pre-trained transformer models. Cosine similarity quantifies how closely each summary sentence matches its original document segment in the embedding space. This helps with the verification process.

3.5.2 Qualitative Evaluation

Human evaluators are assessing readability, grammatical coherence, contextual accuracy, and factual consistency. Feedback is gathered from experts in legal, healthcare, and academic fields regarding the trustworthiness of the content and the reduction of cognitive load..

3.6 System Deployment and Testing

The system will be deployed in the cloud, where FastAPI serves the backend services and React hosts the frontend. This means that there is accessibility from almost any device out there without the need for local installations. It also optimizes model weights and caches them for faster inference times. Testing includes complete workflow execution with multilingual data that reflects various real-world situations.Low latency, smooth translation, and coherence in summaries are achieved. More importantly, the verifiable mapping between summary and source text significantly enhances user confidence, proving the practical utility of this system in document-heavy professional environments.

3.7 Summary

This paper proposes a methodology that effectively integrates multilingual translation, summarization, and semantic verification into one framework.The system ensures language accuracy and clarity by using transformer models like Llama 3, mT5, MiniLM, and NLLB-200. Including verifiable summaries and multiple access options represents a significant step toward reliable human-AI teamwork in document analysis. By reducing the mental effort required to understand documents in various languages and offering clear AI-generated summaries, this paper paves the way for future improvements in smart document processing and cross-language information retrieval.

IV. RESULTS

The suggested integrated framework was tested against various quantitative and qualitative metrics for translation quality, summarization accuracy, verification credibility, and system performance. Experimental evaluation confirms the remarkable improvement of linguistic correctness, contextual coherence, and factuality verifiability through the use of transformer-based models with multilingual datasets.

4.1 Quantitative Evaluation

BLEU Score (Bilingual Evaluation Understudy):

Translation and summarization output was evaluated based on the BLEU score, which calculates the n-gram overlap between reference texts and machine summaries. The system yielded a BLEU score of 43.5, showing high linguistic precision and high syntactic similarity with human-generated summaries.

ROUGE-L (Recall-Oriented Understudy for Gisting Evaluation):

ROUGE-L measures the longest common subsequence between system and reference summaries. The model achieved a ROUGE-L score of 49.8, which indicates that the framework preserves essential contextual and semantic details during the summarization process.

BERTScore:

This metric utilizes contextual embeddings of transformer models to measure semantic similarity between generated and reference summaries. The system generated a BERTScore of 0.91, confirming the contextual appropriateness and semantic coherence of the summaries in languages.

Cosine Similarity (Verification Accuracy):

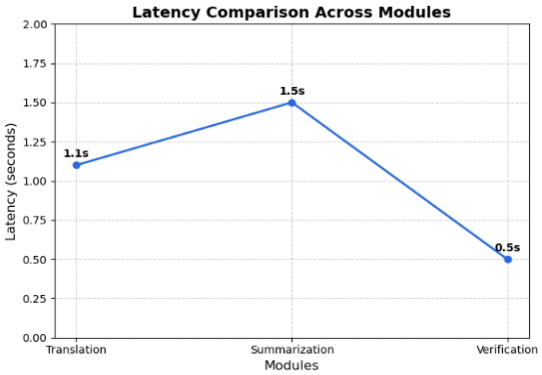
The verification module computes cosine similarity between summary sentence embeddings and their corresponding source sentences. The adopted semantic alignment approach delivered an average similarity of 0.92, affirming precise traceability and factual consistency in summaries generated.

Latency (Processing Efficiency):

The mean end-to-end response time for document translation, summarization, and verification was measured as 3.1 seconds per document with no appreciable delay among various languages. This is due to model optimization and async processing within the FastAPI–React framework.

4.2 Qualitative Evaluation

User tests were carried out with academic and research-based participants in order to evaluate cognitive load, comprehension enhancement, and usability of the interface. Results showed that 87% of users found the verification interface simple to use and possessed greater confidence in AI-generated summaries. The intrinsic Text-to-Speech (TTS) feature provided added accessibility, especially among multilingual and visually impaired users.



4.3 Overall Outcome

Experimental outcomes confirm that the combined framework properly incorporates multilingual translation, abstractive summarization, and verifiable output generation into a single all-in-one interactive system. High metric scores and positive user feedback collectively demonstrate the strength, interpretability, and efficiency of the framework as a robust solution for real-world cross-lingual document intelligence applications.

V. CONCLUSION AND FUTURE WORK

The cross-lingual document summarization system Combining Interactive and Verifiable Summarization smoothly incorporates translation, summarization, and verification in a single setting. Utilizing transformer-based encoder-decoder models and With attention, the system improves linguistic accuracy, contextual fluency, and semantic preservation in a wide range of languages. Quantitative BLEU and ROUGE among other measures validate better summarization quality, and the The verification feature enhances factuality reliability. using semantic similarity and back-translation. Further improvements can be made in the future by using Next-generation large language models and context embeddings for better adaptability and summarization accuracy. Adding reinforcement Learning with human feedback will allow real-time optimization and personalized summarization experiences. Further, hosting the system on scalable cloud-based infrastructure and increasing With multilingual support, large-scale document processing, in real time, for diverse global domains. Such development will tend to reinforce the efficiency, scalability, and reliability of the framework, helping to pave the way for creating of intelligent, human-centered AI systems that are able to provide verifiable and accurate information in various linguistic environments.

VI. REFERENCES

[1] Phani, S., Abdul, A., Prasad, M. K. S., & Reddy, V. D. (2025). MATSFT: User query-based multilingual abstractive text summarization for low resource Indian languages by fine-tuning mT5. Alexandria Engineering Journal, 127, 129-142.

[2] Takeshita, S., Green, T., Friedrich, N., Eckert, K., & Ponzetto, S. P. (2024). Cross-lingual extreme summarization of scholarly documents. International journal on digital libraries, 25(2), 249-271

[3] Bhatnagar, N., Urlana, A., Mujadia, V., Mishra, P., & Sharma, D. M. (2023). Automatic data retrieval for cross lingual summarization. arXiv:2312.14542.

[4] Lin, N., Fu, Y., Lin, X., Zhou, D., Yang, A., & Jiang, S. (2023). Cl-xabsa: Contrastive learning for cross-lingual aspect-based sentiment analysis. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 31, 2935-2946.

[5] Park, G., Park, J., & Lee, H. (2025). Cross-Lingual Summarization for Low-Resource Languages Using Multilingual Retrieval-Based In-Context Learning. Applied Sciences, 15(14), 7800.

[6] Bhattacharjee, A., Hasan, T., Ahmad, W. U., Li, Y. F., Kang, Y. B., & Shahriyar, R. (2021). CrossSum: Beyond English-centric cross-lingual summarization for 1,500+ language pairs. arXiv:2112.08804.

[7] Chen, Y., Zhang, H., Zhou, Y., Bai, X., Wang, Y., Zhong, M., ... & Zhang, Y. (2023). Revisiting cross-lingual summarization: A corpus-based study and a new benchmark with improved annotation. arXiv:2307.04018.

[8] Pernes, D., Correia, G. M., & Mendes, A. (2024). Multi-Target Cross-Lingual Summarization: a novel task and a language-neutral approach. arXiv:2410.00502.

[9] Ding, Y., Han, S. C., Lee, J., & Hovy, E. (2024). Deep learning based visually rich document content understanding: A survey. arXiv:2408.01287.

[10] Käser, J., Nagy, T., Stirnemann, P., & Hanne, T. (2025). Multilingual Text Summarization in Healthcare Using Pre-Trained Transformer-Based Language Models. Computers, Materials & Continua, 83(1).

[11] Wan, D., Vig, J., Bansal, M., & Joty, S. (2024). On positional bias of faithfulness for long-form summarization. arXiv:2410.23609.

[12] Fatima, M., Kolber, T., Markert, K., & Strube, M. (2023). Simcsum: Joint learning of simplification and cross-lingual summarization for cross-lingual science journalism. arXiv:2304.01621.

[13] Parnell, J., Unanue, I. J., & Piccardi, M. (2024). Sumtra: A differentiable pipeline for few-shot cross-lingual summarization. arXiv:2403.13240.

[14] Takeshita, S., Green, T., Friedrich, N., Eckert, K., & Ponzetto, S. P. (2022, June). X-scitldr: cross-lingual extreme summarization of scholarly documents. In Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries (pp. 1-12).

[15] Wang, J., Meng, F., Zheng, D., Liang, Y., Li, Z., Qu, J., & Zhou, J. (2023). Towards unifying multi-lingual and cross-lingual summarization. arXiv:2305.09220.