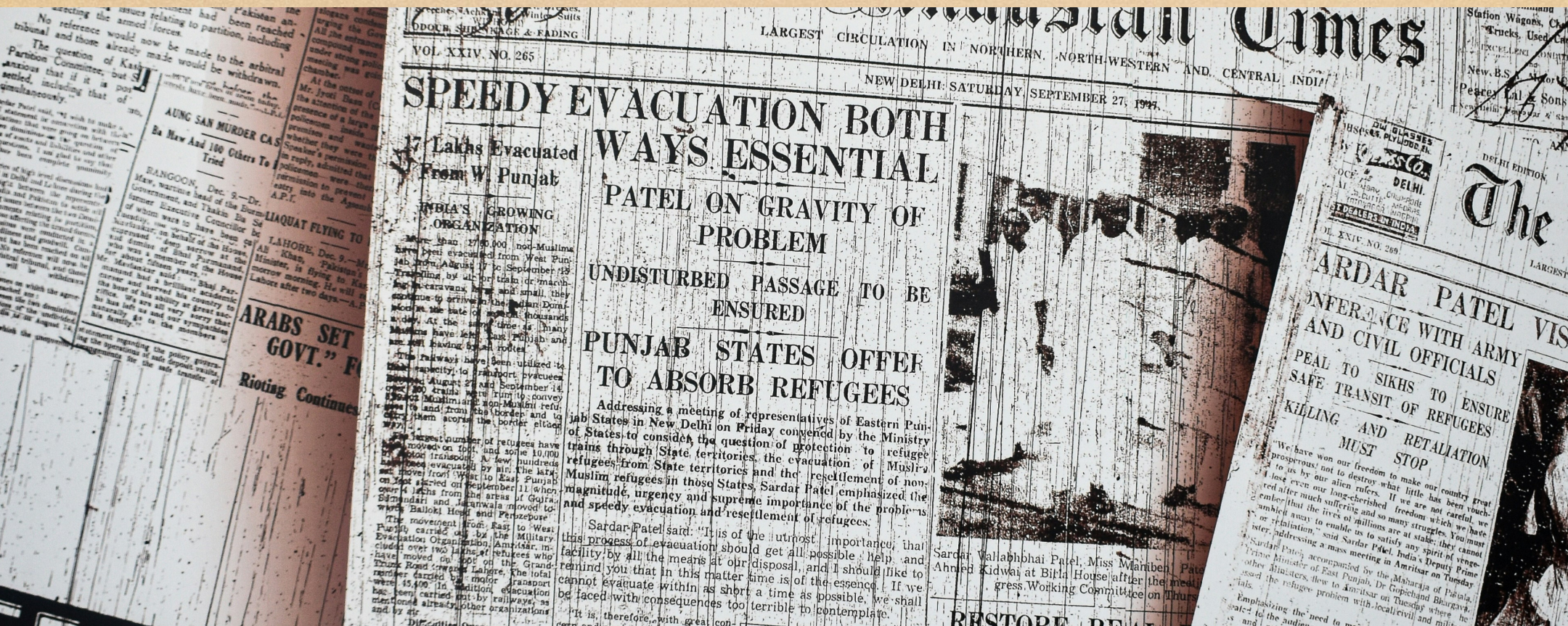
TEAM DETAILS AND PROBLEM STATEMENT

- Problem Statement : Global Article Summarization and Link Analysis Platform
- Team Name : Resc.ai
- Team Leader Details (Name, Phone Number, Email) : Devesh mishra,+916306962393, mishradevesh5518@gmail.com
- Institution Name : Bennett university
- Course Enrolled : Btech cse

Problem statement

- Develop an interactive platform using advanced language processing to filter and summarize articles based on user inputs.
- Users struggle to extract insights from the vast amount of available information, necessitating concise, relevant summaries.
- The platform should be customizable and analyze relationships between articles to enhance understanding.



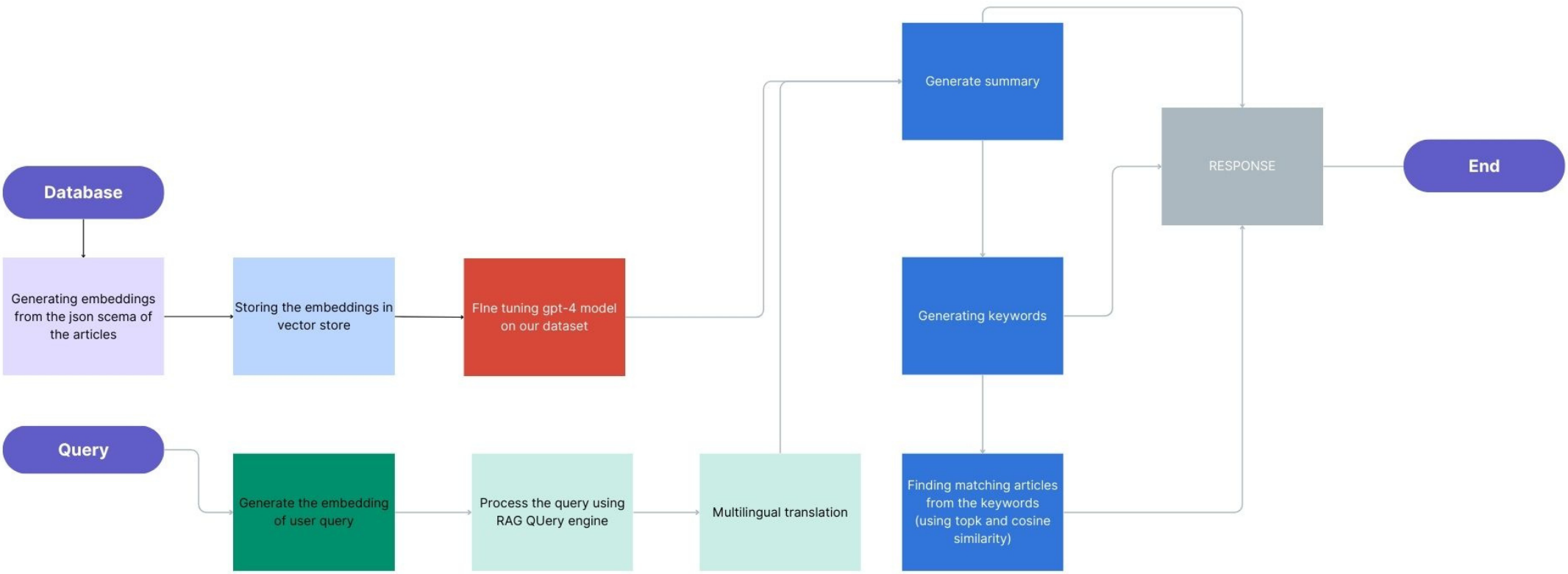
Outline the goals and objective of your approach and what you are aiming to achieve.

1. **Develop a newspaper article summarizer using a custom dataset and fine-tuned models.**
2. **Integrate LlamaIndex for efficient article indexing and retrieval.**
3. **Utilize GPT-4 embeddings to enhance the summarization process.**
4. **Ensure the summarizer produces concise and relevant summaries across diverse topics.**
5. **Evaluate summarizer performance based on accuracy, coherence, and relevance metrics.**
6. **Implement a user-friendly interface for inputting queries and accessing article summaries.**
7. **Enable the summarizer to suggest related articles based on summary content, promoting user engagement.**

Describe your Methodology

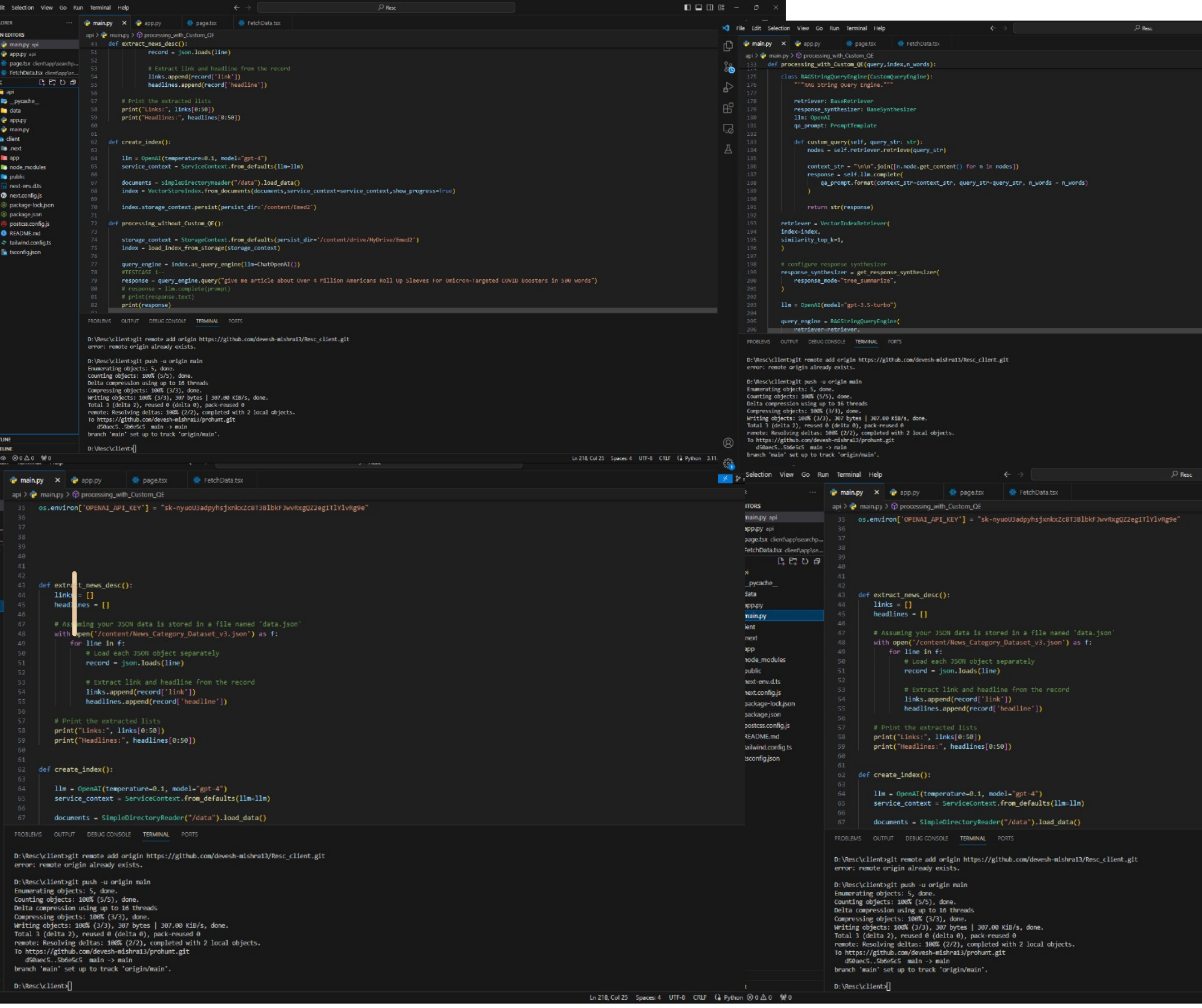
1. **Data Collection and Preprocessing: Gather diverse newspaper articles and preprocess them to remove noise and standardize formatting.**
2. **Model Selection: Choose GPT-4 as the language model for fine-tuning due to its advanced language processing capabilities.**
3. **Fine-tuning: Adapt the selected language model to the task of article summarization taking reference from news-article-dataset.json then creating our own dataset of 10k+ news articles for better description**
4. **LlamaIndex Integration: Integrate LlamaIndex for efficient indexing and recurrsive retreival with response synthesizer using topk similarity for better finetune result.**
5. **Embedding Generation: Generate embeddings for articles using GPT-4-0613 7B model, enhancing summarization quality and semantic understanding.**
6. **Evaluation and User Interface: Assess summarizer performance using relevant metrics and develop a user-friendly interface for easy access to multilingual article summaries and related suggestions.**

Relevant image related to your idea.



Mention your technology stack

1. **LLAMA Index**
    - **VectorStoreIndex**
    - **CustomQueryEngine**
2. **OpenAI GPT-4**
3. **Langchain**
4. **Google Translator API**
5. **deep_translator**
6. **rouge_score**
    - **NLTK (Natural Language Toolkit)**



Github backend link with summaries - **https://github.com/devesh-mishra13/Resc**

Github frontend link - **https://github.com/devesh-mishra13/Resc_client**

Describe your use cases.

1. Article Summarization
2. Keyword Extraction
3. Related Article Suggestions
4. Customized Summarization Length
5. Language Support
6. Interactive User Interface
7. Educational and Research Purposes

Future Work.

1. Incorporating User Feedback Mechanisms for Improved Summaries
2. Integrating Semantic Understanding for Better Keyword Extraction
3. Enhancing Link Analysis for More Relevant Article Suggestions
4. Expanding Language Support to Include More Languages
5. Developing Mobile Applications for Accessibility
6. Exploring Integration with Voice Assistants for Hands-free Interaction