Enhanced Accessibility and Comprehension of Scientific Literature through AI

Team Members
1. Gowtham Medisetti - 11635487
2. Anisha Peddireddy - 11649037
3. Mukunda Krishna Ramisetti -11647459
4. Sriharshini Saraswathi - 11655475

Github repository:
https://github.com/Gowtham-924/Enhanced-Accessibility-and-Comprehension-of-Scientific-Literature-through-AI-

Motivation
Navigating the vast ocean of scientific literature can be daunting. Researchers may spend hours searching for relevant papers, only to face the additional challenge of deciphering dense and complex articles. This project addresses the need for a streamlined approach to access and understand scientific texts, facilitating quicker insights and broader comprehension.

Significance
The significance of this project lies in its potential to:
Simplify Research: To increase the effectiveness of their work, a researcher studying renewable energy can easily find relevant studies on the newest solar cell technology.
Educate and Inform: Students' learning journeys are supported by easily accessible explanations of intricate biochemical processes that have been simplified.

3. Objectives

- Goal 1: Create a model that, for instance, can accurately classify papers into physics and chemical categories.
- Goal 2: Develop a method for summarizing information that generates concise, understandable summaries. An overview of a study on a novel cancer treatment, for instance, might emphasize the course of treatment, outcomes of tests, and any possible ramifications.
- Goal 3: Attain a respectable accuracy rate of consumer satisfaction. pointing out that the classifications and summaries greatly support their research or study activities.

4. Features

i.AI-Powered Classification: Putting a paper on "Neural Networks in Autonomous Vehicles" that has been uploaded into the Computer Science category automatically.

ii.Summaries in Multiple Modes: creating a key points summary for a paper on "The Effects of Climate Change on Coral Reefs."

iii.Interactive Platform: Users can rate the usefulness of a summary by saying, for example, whether or not an explanation of quantum computing concepts was simple to understand.

5. Dataset

1.Text Classification: Utilizing a subset of the arXiv Dataset comprising 100,000
Dataset 1: arXiv Subset
Size: 100,000 scientific articles
Type: Text
Source: arXiv, a collection of research papers
Preprocessing: Cleaning up the text by removing non-essential stuff like headers, punctuation, and normalizing text (making it consistent, like changing "cats" to "cat").

2.Text Summarization: Leveraging 50,000 papers from CORD-19,
Dataset 2: CORD-19 Subset
Size: 50,000 papers
Type: Mix of text and images
Source: CORD-19, a collection of COVID-19 related papers
Preprocessing: Cleaning, organizing, and simplifying the text.

6. Visualization

This project's workflow is a step-by-step procedure designed to accomplish its main goals. Starting with text categorization, scientific papers are categorized by the algorithm into relevant domains like chemistry or physics. The material is next summarized, in which the most important ideas and conclusions are condensed into brief and clear explanations. In an effort to achieve a high degree of user satisfaction, the system lastly requests comments and ratings from users in order to assess the efficacy of the created summaries. The project's objectives of precise classification, succinct summarizing, and user pleasure are strongly aligned with this methodology, which guarantees a streamlined approach to improving accessibility and comprehension of scientific literature. Workflow explained in the below figure1.

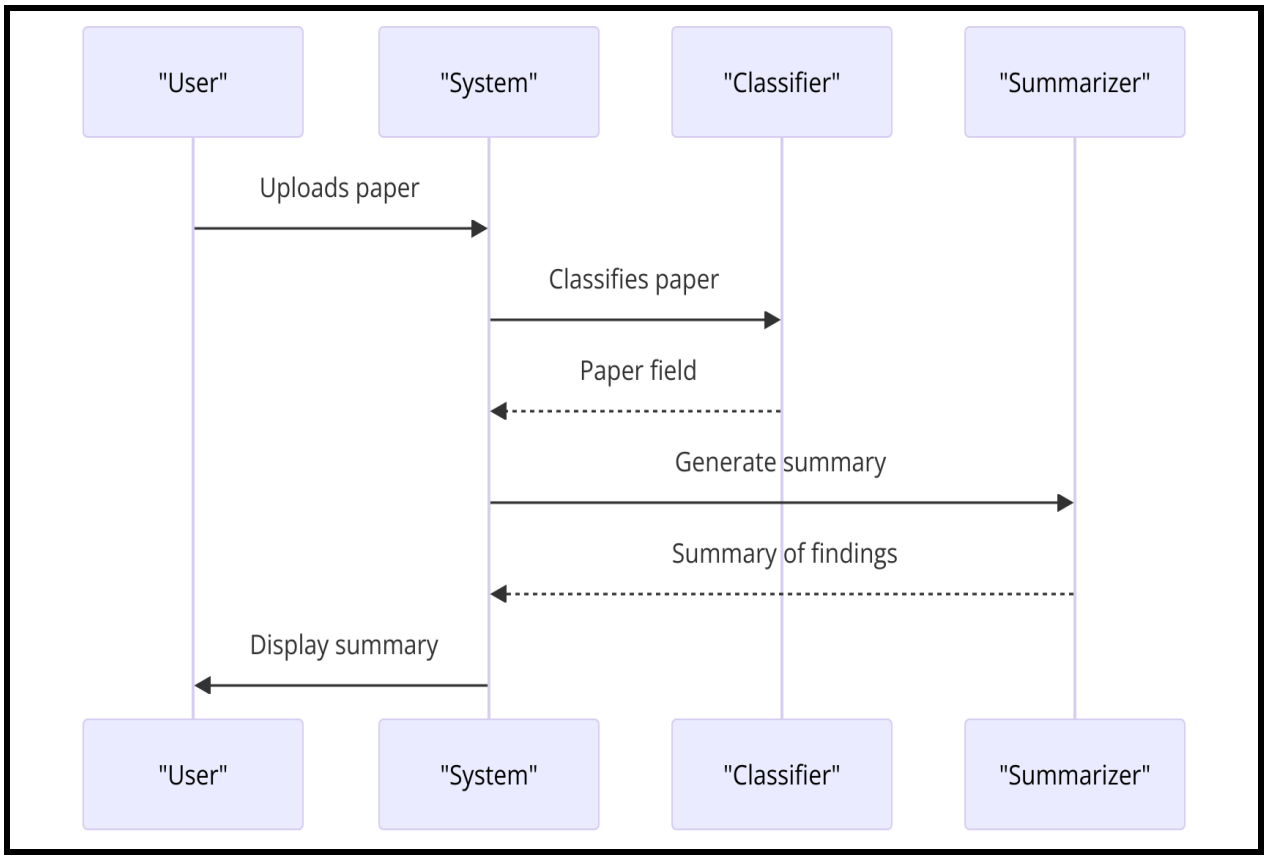

Fig 1: Work flow diagram for Enhanced Accessibility and Comprehension of Scientific Literature through AI