

DocSense: A Natural Language-Powered Search and Summarization Framework

Saurav Verma¹ Siddhanth Sharma², Yash Shrivastava³, Rutuja Jadhav⁴

Mukesh Patel School of Technology Management & Engineering, SVKM's NMIMS,

Mumbai, India^{1,2,3,4}

Article History:

Received: 12-01-2025

Revised: 15-02-2025

Accepted: 01-03-2025

Abstract: In the digital era, decision-making and knowledge sharing depend on the ability to acquire and derive relevant insights from a variety of information sources. When it comes to handling complicated papers, CSV files, and multimedia content, traditional chatbots are limited, which frequently results in a lack of usefulness and accuracy. Because of these restrictions, sophisticated AI-powered solutions like DocSense must be developed. With its innovative AI-powered Chatbot, DocSense, customers may submit documents, CSV files, and YouTube videos for smart analysis—a feature that addresses the drawbacks of other chatbots. Present-day chatbots frequently have trouble summarising documents, which makes it difficult for users to quickly and effectively extract important information. DocSense gets around this problem by using cutting edge natural language processing methods to deliver extremely accurate document summarization. The DocSense Enhanced Natural-Language Search Engine is one of its main innovations. Conventional chatbots generally rely on keyword-based searches, which may produce results that are insufficient or irrelevant. DocSense, on the other hand, uses sophisticated language models to comprehend user inquiries in a context-aware way. This allows users to get accurate insights from their submitted documents, CSV data, and YouTube videos and ask queries in natural language. Users can upload a variety of document formats with DocSense, such as Word documents, Excel spreadsheets, PDFs, and more. These files are processed quickly by the AI-powered Chatbot, which then creates clear, coherent summaries that include the most important details. Furthermore, capable of transcribing and summarising YouTube videos, DocSense is a priceless resource for scholars, content providers, and decision-makers who want in-depth understanding from a variety of multimedia sources

Keywords: Lorem, ipsum, dolor, sit.

1. INTRODUCTION

In a time where there is a never-ending digital information vault, being able to quickly glean insightful information from a variety of sources has become essential. Conventional chatbots are useful for facilitating human-computer communication, but they are not always up to the task of managing complicated documents, CSV files, or multimedia information. These restrictions include the incapacity to precisely handle documents and the deficiency of powerful natural language search tools. In DocSense: Document Summarization with Enhanced Natural-Language Search Engine in this area. With the goal of addressing the issues with existing Chatbot solutions, DocSense is a revolutionary AI-powered Chatbot. The inadequate summary of documents, a crucial component of knowledge management and information retrieval, is one of the most common problems. Current chatbots

frequently fail to deliver succinct, insightful document summaries, forcing users to deal with long messages and decreased accuracy. DocSense provides unmatched accuracy and efficiency in document summarising by utilising state-of-the-art natural language processing algorithms. DocSense is known for its innovative Enhanced Natural-Language Search Engine. Traditional chatbots mostly rely on keyword-based searches, which is a methodology that frequently yields vague or insufficient responses. DocSense, on the other hand, adopts a transformative strategy by utilising sophisticated language models to understand user inquiries in their particular circumstances. This enables users to interact naturally and conversationally with DocSense, asking questions in simple terms and getting detailed insights from uploaded documents, CSV files, and even YouTube videos. One feature that makes DocSense a very useful tool is its ability to process a large variety of document types, including Word, Excel, and PDF files. DocSense's capabilities extend beyond text; it can also be used to analyse YouTube videos, transcribe their content, and produce concise summaries. Those who create content, conduct research, and make decisions who want to access the abundance of information found in multimedia sources will find this feature especially helpful. By tackling the underlying issues that other solutions to the field of AI-powered chatbots were unable to resolve, DocSense marks a significant advancement in the field.

DocSense's advanced Enhanced Natural-Language Search Engine, proficiency in document summarising, and versatility in handling various document types enable users to quickly and easily derive valuable insights. DocSense is a cutting-edge tool that has the potential to completely change how people and organisations use AI for document analysis and information retrieval in the age of data-driven decision-making. The detail about the features and capabilities that make DocSense a valuable tool for today's information seekers in the pages that follow. The goal is to motivate people and organisations to use DocSense for a variety of important goals. By allowing users to quickly obtain relevant information from papers and multimedia content, it promotes efficiency improvement by helping users save time and effort. Additionally, it facilitates well-informed decision-making by utilising the accuracy and comprehensiveness of DocSense's summarising features provide decision-makers a better comprehension of their data. Additionally, it promotes information sharing by enabling professionals, academics, and content creators to glean priceless insights from multimedia sources

2. LITERATURE SURVEY

The constantly evolving fields of artificial intelligence (AI) and natural language processing have spawned a plethora of creative solutions designed to glean insightful data from a variety of sources. In this, it important to explore the underlying technologies, difficulties, and developments that form the basis of DocSense, an advanced AI-powered Chatbot, in this thorough literature review[1]. Document summarising, multimedia content analysis from YouTube videos, data processing and visualisation from CSV files, and other state-of-the-art features are all provided by DocSense[2]. In this work, the use of numerical citations to cite important research publications throughout this evaluation in order to give a thorough examination of the environment in which DocSense functions. Over time, chatbots have undergone an incredible amount of change. Chatbots, which were formerly limited to basic communication, have developed into intelligent conversational assistants with a wide range of uses

[1]. In begin our analysis with a brief history of Chatbot development, which helps to illustrate the increasing attention being paid to these conversational agents around the world [1]. This historical viewpoint is essential to comprehending the path that gave rise to platforms such as DocSense. It is also investigate the reasons behind the application of chatbots in a variety of fields, such as marketing, teaching, and medical [1]. This historical viewpoint provides context for DocSense's capabilities, which include document summarization and insight extraction from a variety of multimedia sources [1][2].

The ability of DocSense to automate document summarising is a crucial component of its capabilities, providing customers with significant time and resource savings [2]. This feature is consistent with study findings from a pertinent paper that discusses the difficulties of creating and maintaining chatbots by hand [3]. The study highlights the growing need for chatbots that can effectively address inquiries, enable self-service, and present goods and services, particularly in the business context [3]. Additionally, the paper presents a novel method that uses web API standards to automatically create chatbots [3]. This methodology is consistent with DocSense's capacity to handle structured data like CSV files and deliver insights via data visualization [2]. In the modern world of technology, chatbots have developed into conversational agents that can efficiently engage consumers and offer support via interactive text messaging [3]. Nonetheless, the creation and coding of chatbots presents a unique set of present-day as well as prospective difficulties [3]. It is essential to comprehend these programming issues because DocSense works in the field of advanced artificial intelligence and natural language processing [3]. A distinct set of challenges is presented by the dynamic nature of chatbots, which combine artificial intelligence, natural language processing, programming, and conversational services [3]. This realisation is essential to DocSense's ongoing evolution and adaption as it traverses the constantly changing landscape of technology and user requirements. DocSense's core feature includes sentiment analysis and data mining, which are especially important when working with multimedia content [4]. A study has looked into the creation, instruction, and enhancement of "Skillbot," a Chatbot with sentiment analysis and data mining capabilities [4]. The study comprised data scraping from the Gov. UK website, intent training for the Skillbot model, and in-depth conversation analysis with users [4]. An important factor in teaching chatbots to respond to users efficiently was sentiment analysis [4]. This study supports DocSense's methodology by emphasising the combination of many technologies, including sentiment analysis models, natural language processing, and visualisation strategies [4]. The end objective in both situations is to enable chatbots to analyse data and user discussions in order to provide better services to users.

The launch of Llama 2 represents a significant milestone in the field of large language models (LLMs), as described in a recent research publication [5]. A set of refined LLMs with scales varying from billions to tens of billions of parameters are presented in Llama 2[5]. These models, called Llama 2-Chat, outperform open-source chat models on a number of benchmarks and are optimised for dialogue use cases [5]. The capabilities of chatbots and AI-powered conversational agents have advanced significantly with the release of Llama 2[5]. In order to leverage this breakthrough, DocSense integrates Llama 2 to comprehend user inquiries and offer context-aware responses [5]. Users may

now interact with DocSense using natural language thanks to this connection, which is a big improvement over conventional chatbots that rely on keyword-based searches [5].

An extensive analysis of the history, uses, difficulties, and potential applications of AI-driven conversational agents is provided by the review of ChatGPT [8]. This thorough analysis includes a look of ChatGPT's many applications, which span industries like healthcare and education [8]. This study's findings are consistent with DocSense's mission to offer users insightful information in a variety of fields [8]. The review highlights the limitations, ethical issues, and potential difficulties that AI chatbots may present, highlighting the significance of responsible development [8]. This understanding guides DocSense's strategy for addressing the difficulties and moral dilemmas raised by AI chatbots while maximising their enormous potential.

In DocSense's responsiveness and performance are improved by the incorporation of Llama 2, an improved big language model [5]. DocSense makes advantage of this foundation to give consumers more accurate, context-aware responses. Llama 2-Chat models have proven to perform better than open-source chat models on a number of benchmarks [5]. DocSense differs from conventional chatbots in that it can comprehend user inquiries in plain language and provide accurate insights from a variety of sources [5]. However, it's important to recognise that, as the research [5] makes clear, there are trade-offs associated with the growth of LLMs. Although Llama 2 offers significant enhancements in terms of performance and capability, processing time may be affected by the models' billions to tens of billions of parameters [5]. DocSense chose the Llama 2 model with seven billion parameters in order to strike a compromise between efficiency and accuracy [5]. A 30-billion-parameter model might increase accuracy, but it might also cause noticeable processing delays. In order to give users a responsive and effective experience, DocSense thereby strikes a balance [5].

DocSense has many benefits, but it is not without drawbacks [2]. The possibility of losing context and subtlety when condensing long papers is one of the fundamental issues of document summarization [2]. Even though DocSense is excellent at producing succinct and logical summaries, subtle nuances can get left out [2]. Additionally, elements like background noise and audio quality may have an impact on how accurate multimedia content analysis is, especially when it comes to YouTube videos [2]. This drawback emphasises the necessity of ongoing development for AI-driven solutions such as DocSense [2]. To sum up, DocSense is an innovative development in the field of artificial intelligence chatbots that enables users to get valuable insights from a variety of sources, including documents, CSV files, and multimedia information such as YouTube videos [2]. Because of its incorporation of Llama 2 and sophisticated NLP methods, it is positioned as an effective and context-aware tool for researchers, content producers, and decision makers [5]. It balances accuracy and performance requirements while deftly navigating the terrain of complicated texts and multimedia content [5]. DocSense has limitations, nevertheless, just like any other technology, especially when it comes to multimedia content analysis and sophisticated document summarization [2]. However, its dedication to ongoing enhancement and future research and development portends far more sophisticated AI solutions [2]. With an eye towards the future, DocSense plans to provide more functionality to serve a larger user base. It seeks to provide a general Chatbot feature that will allow users to converse in natural language

about a variety of topics [5]. DocSense will become an even more flexible information retrieval tool as a result of its progress, which will also improve the user experience [5]. Furthermore, the incorporation of URL-based article summary is imminent [5]. Researchers and content consumers will find this capability to be a priceless resource as it will enable users to extract important information from online articles [5].

3. METHODOLOGY AND IMPLEMENTATION

3.1 Model Selection

Large Language Models, or LLMs for short, are a revolutionary development in artificial intelligence and natural language processing (NLP). These models are enormous, with billions of parameters on average, and are trained using enormous volumes of text data from the internet. They are a part of the larger class of machine learning models called deep neural networks. LLMs also facilitate the extraction of knowledge from large textual databases. Their ability to sort through large amounts of text and retrieve insightful information is extremely helpful to researchers and analysts, as it would be difficult and time-consuming for people to accomplish by hand. Data privacy is a further issue. Due of their large memory, LLMs may unintentionally memorise and reveal private information found in their training datasets. One of the most important aspects of developing and implementing LLM is maintaining data privacy and protection.

LLMs are expected to continue to improve and have new uses in the future. The fact that these models are still getting bigger and more capable is one of the most obvious trends. This advancement pushes the limits of natural language processing, enabling even more precise and contextually aware text generation and understanding. In summary, large language models are changing the field of natural language processing and artificial intelligence. Their enormous capacity for text generation and comprehension has far-reaching effects in a variety of fields. But these developments also present privacy, ethical, and environmental issues that demand careful planning and implementation. Future breakthroughs in AI systems are expected to be even more sophisticated, adaptable, and moral, and LLMs will likely continue to be essential to these advancements.

3.2 Methodology

3.2.1 CSV model working overview

DocSense, an advanced AI-powered Chatbot, offers a seamless and user-friendly experience through a well-structured workflow. At the outset, users can choose to load datasets, with options to select from default datasets like Movies, Cars, and Colleges, or to upload custom datasets. These datasets are efficiently stored in a dictionary, ensuring easy access and organization. Upon dataset selection, users have the freedom to opt for two distinct modes: Visualization or Chat. In the Visualization mode, users can enter specific questions related to the dataset and click the "Go" button to receive instant answers. Prior to sending the query, DocSense ensures the validation of Hugging Face API keys and provides a primer on the selected dataset. The chatbots further formats the question and communicates with the Hugging Face API, powered by Code Llama, to retrieve accurate responses.

If applicable, DocSense also generates and displays plots to complement the answers. Conversely, in Chat mode, users can engage in natural language conversations with the chatbots, typing questions and clicking the "Ask" button. Like in the Visualization mode, DocSense takes care of key validation and dataset primers, formatting the chat question for processing. The question is then sent to the Hugging Face API, with the chatbots answer being displayed for user reference. Importantly, DocSense offers the unique functionality of executing answers as code, enhancing its versatility for various use cases. The flow continues, allowing users to repeat these steps for more questions and seamlessly switch between datasets through user-friendly tabs, including options for default datasets and custom uploads. DocSense ensures that users can easily access and navigate through their selected datasets. Furthermore, to give credit to the dataset sources and providers, a footer displaying dataset credits is added, promoting transparency and acknowledgment of data sources.

3.2.2 YouTube model working overview

The workflow of the advanced AI-powered Chatbot, known as DocSense, for YouTube video summarization seamlessly guides users through a structured process to extract meaningful insights from multimedia content. Users initiate the process by inputting a YouTube video URL. Additionally, they have the option to paste their Hugging Face API token for enhanced customization. A "Force Transcription" checkbox allows users to request transcription even if one is already available, giving them full control over the summarization process. To fine-tune the summarization, the Chatbot provides a "Temperature Slider" for setting the generation temperature and a "Summary Length Slider" to determine the desired length of the summary. Users can further customize the generation process with the "Do Sample" checkbox.

Upon clicking the "Summarize" button, DocSense performs a series of operations with precision. First, it verifies the validity of the provided YouTube URL. Subsequently, it processes the video by extracting its content and proceeding to transcribe the audio. Here, DocSense intelligently checks if a transcript is already available and utilizes it whenever possible. In cases where users enable "Force Transcription," the Chatbot takes the additional steps of downloading the audio from the video and transcribing it using the Hugging Face Hub API, ensuring that comprehensive transcripts are available. With the transcript in hand, DocSense employs its powerful Llama code for summarization. The result is a concise and coherent summary of the video's content. The Chatbot not only delivers this summary but also provides valuable information, including the video title, the source of the transcript, and the generated summary. Users can easily repeat this process for more video summarizations.

3.3.3 Flowchart for CSV, Summarizer and Overall of application

This flowchart illustrates a possible flow within an application that involves handling CSV data, utilizing a summarizer, and incorporating a chat module for real-time communication. The application include a chat module where users can communicate in real-time and CSV file is processed by the CSV visualizer to display its data. Users interact with the displayed data shown in Fig.1 & 2.

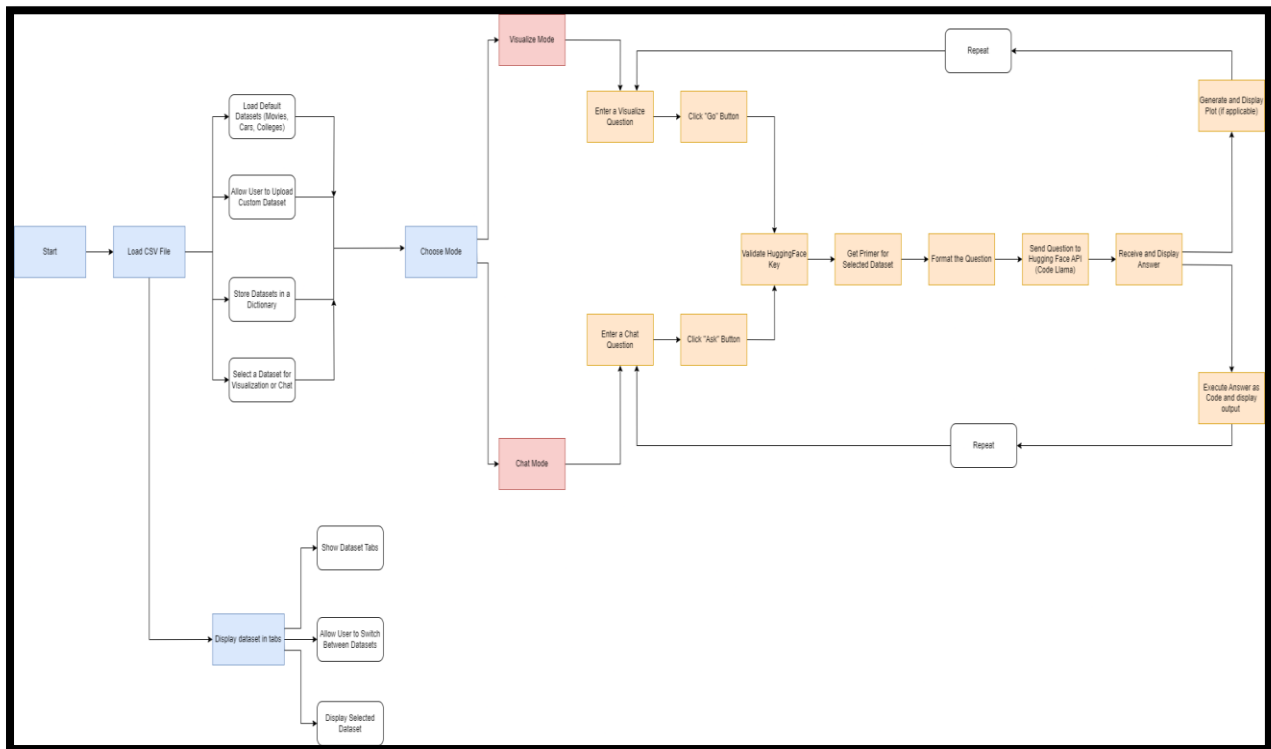


Fig.1: CSV visualizer and Chat module

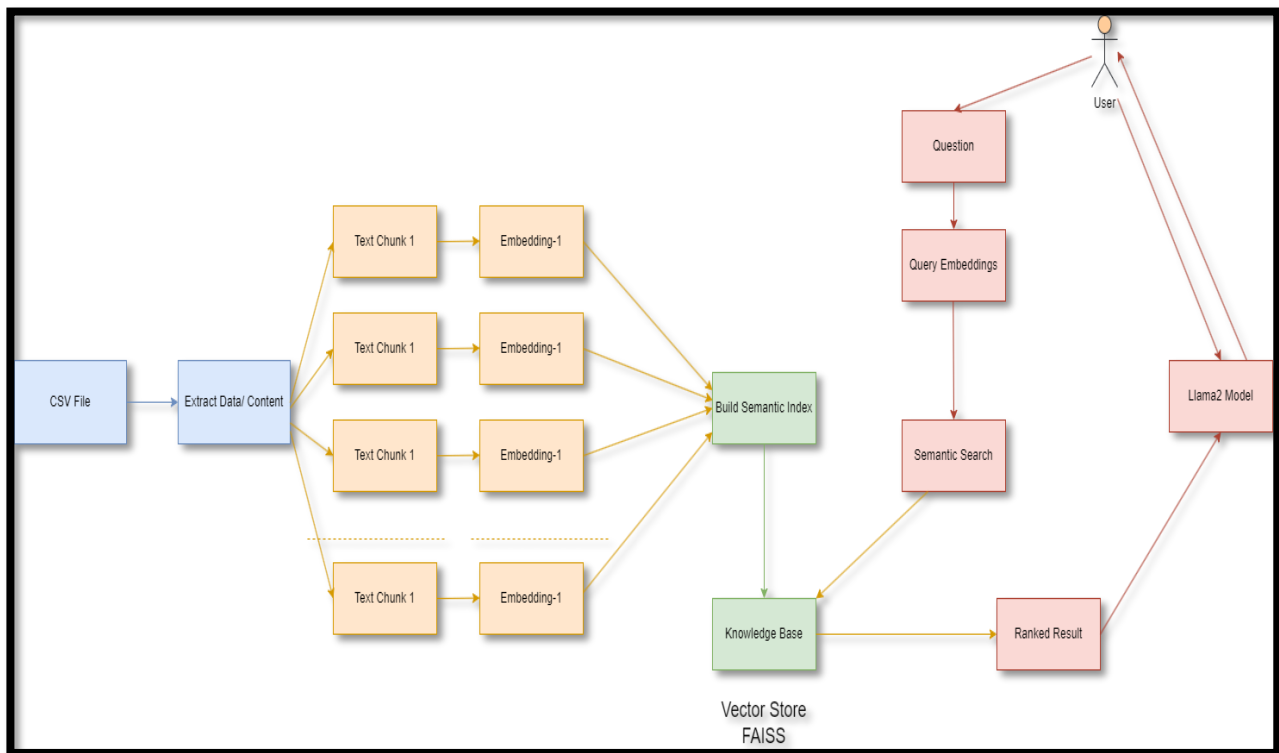


Fig.2: Internal flow of the model

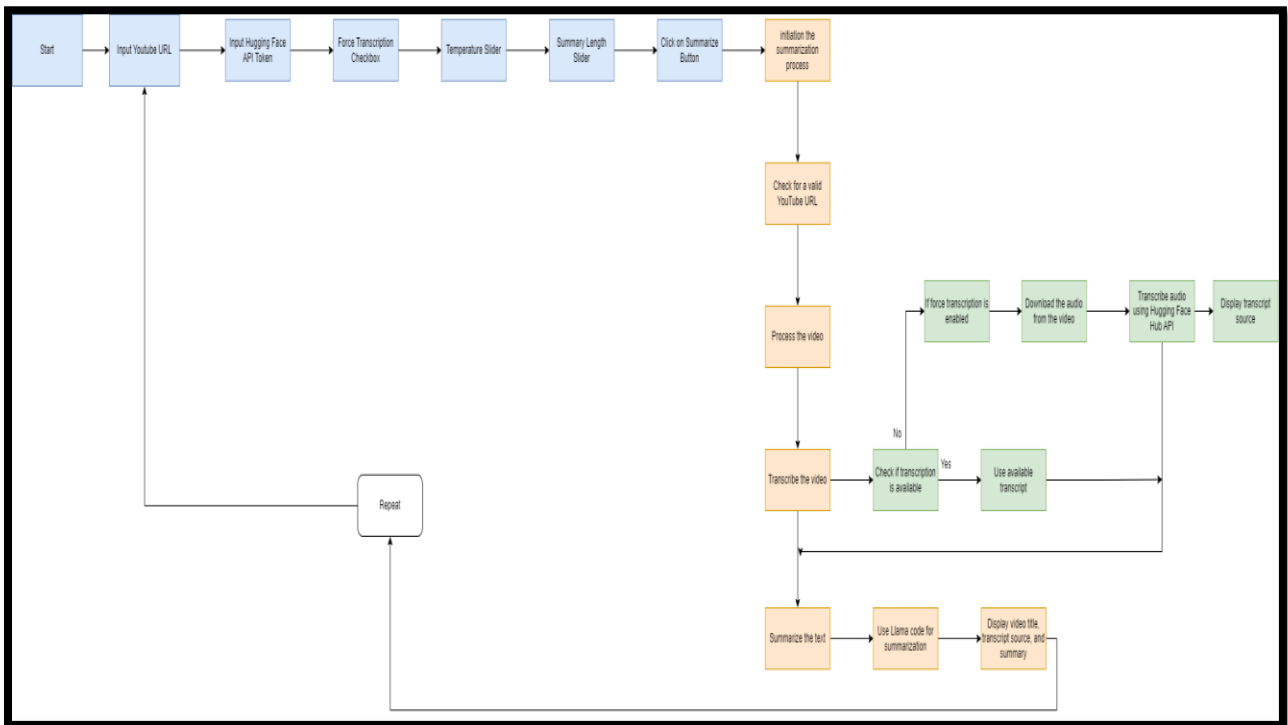


Fig.3: YouTube Summarizer Module

A YouTube summarizer module is a component or tool designed to extract key information from videos hosted on YouTube as shown in Fig.3 & 4.

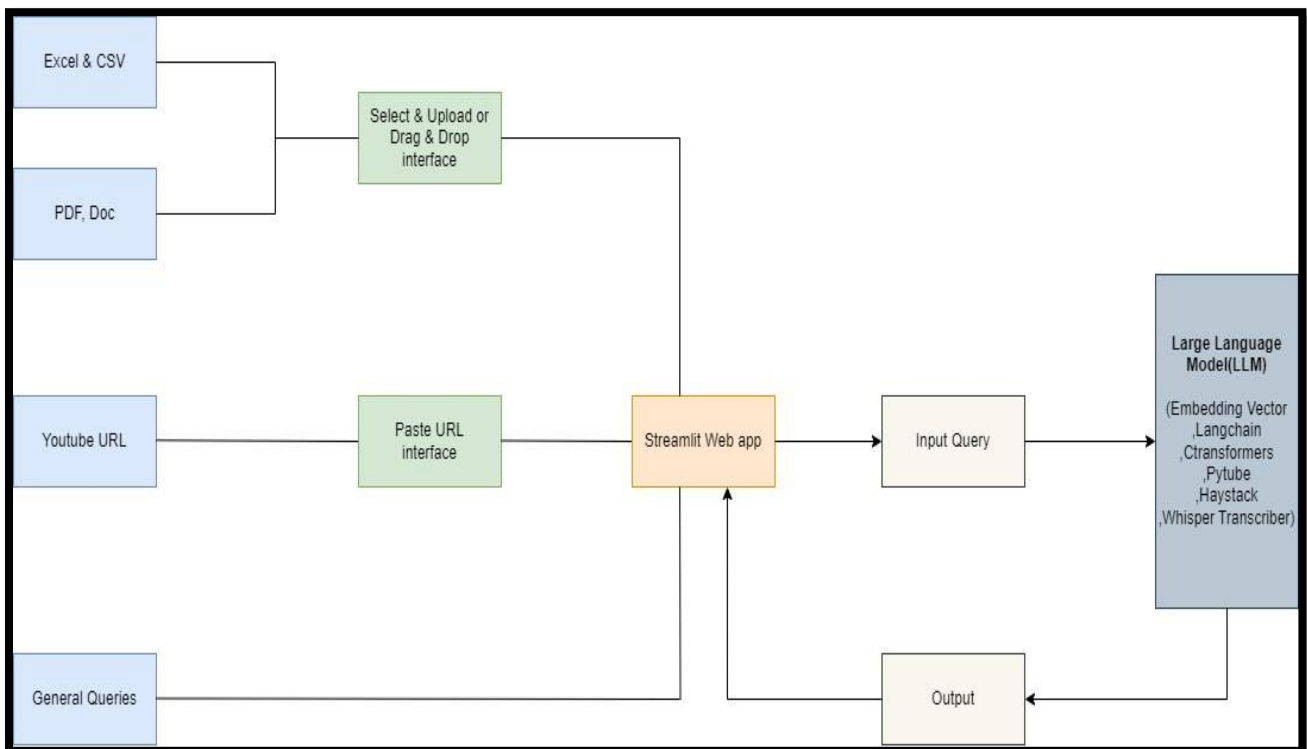


Fig.4: Overall flow of the Application

4. RESULTS AND ANALYSIS

This section discussed about the challenges, notable contributions, results, and analyses of DOCSense's development. Drawing inspiration from innovative methodologies, it is assess how DOCSense overcame these challenges and evaluate its achievements. In this section, shows the results of our system's performance, specifically focusing on the challenge of delayed responses, and provide an analysis of the findings. The performance of our system in producing precise responses to user inquiries constantly showcases the efficacy of our AI models and the capabilities of our natural language processing technology. The system demonstrated its proficiency in comprehending and analysing natural language inputs by providing accurate responses to a diverse array of queries. A further approach employed was the incorporation of external Application Programming Interfaces (APIs) within our system. The selection of these specialised APIs was made with the intention of transferring specific processing responsibilities from the local system, hence potentially enhancing response times. Preliminary examinations of external application programming interfaces (APIs) demonstrated encouraging outcomes in mitigating time lags for particular sorts of queries. In order to achieve a successful integration of external APIs, it was necessary to conduct a comprehensive review and compatibility assessment to guarantee smooth operation within our system architecture.

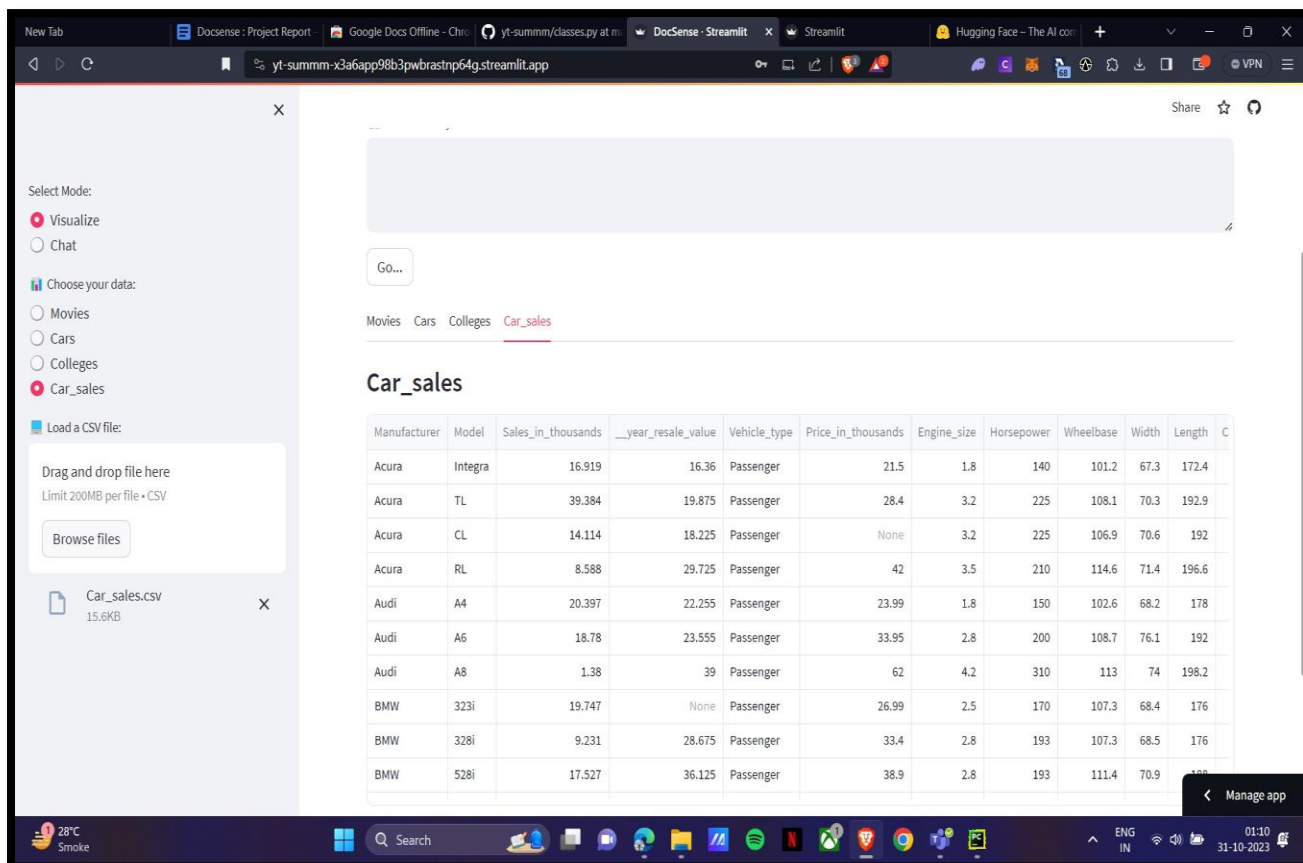


Fig.5: CSV Demonstration

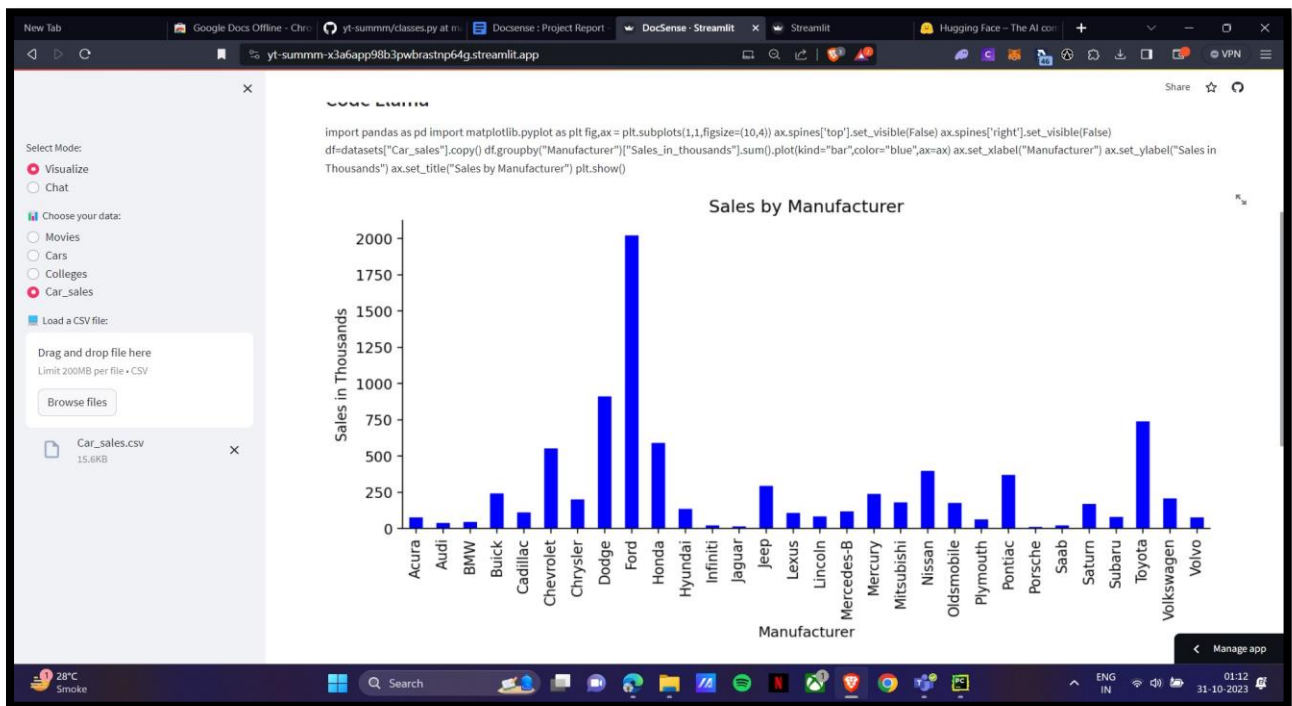


Fig.6: CSV Demonstration

In figure 5 to 9, demonstrate a CSV implementation for various types of data queries and output presented in different visualisations and Fig.10 & 11 show the demonstration of YouTube Summarizer according to DocSense.

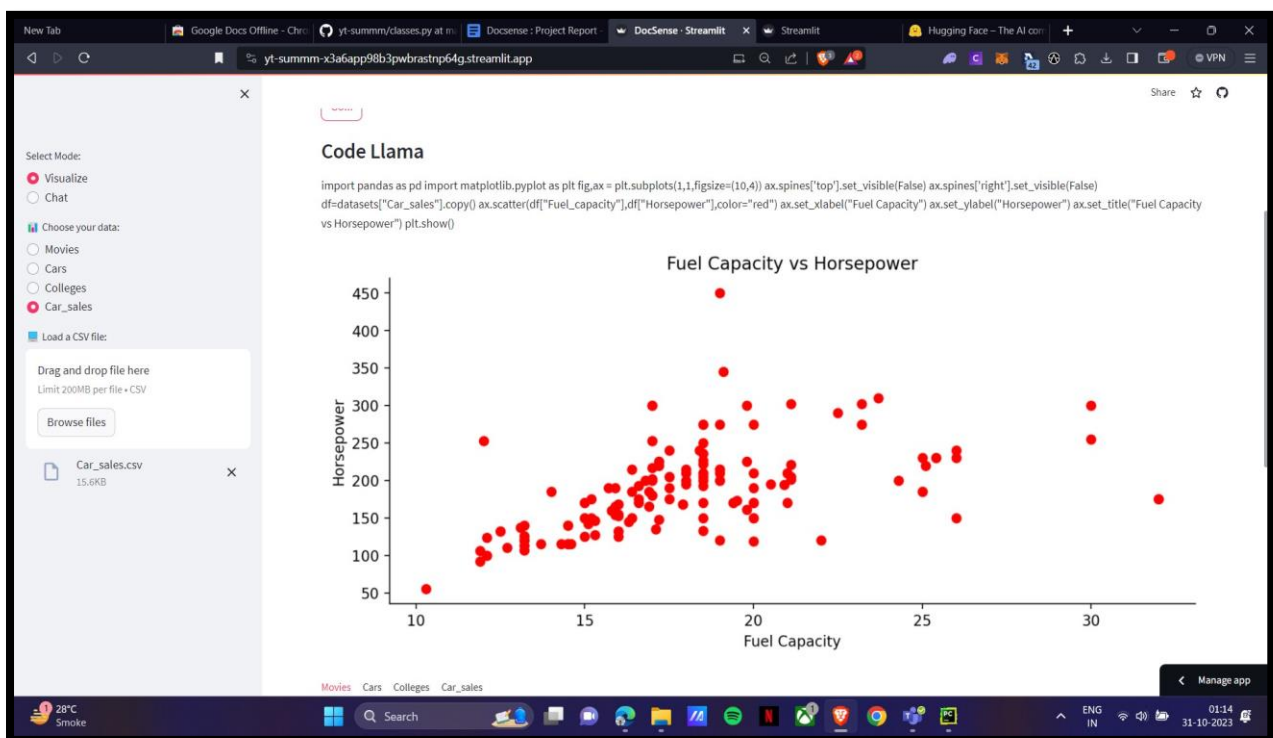


Fig.7: CSV Demonstration

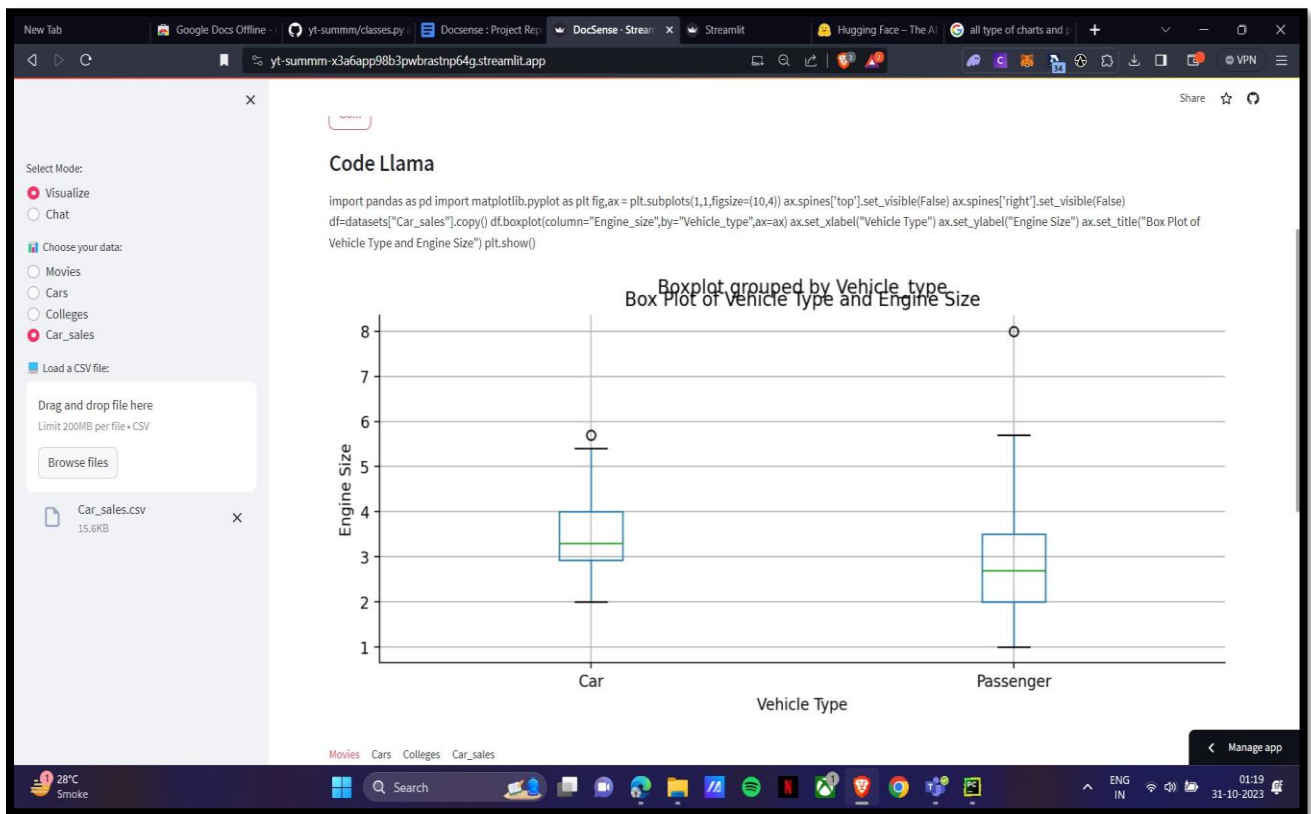


Fig.8: CSV Demonstration

What question do you have about the dataset?

Give me the cars that have fuel capacity more than 15 and sales in thousand is more than 20

Ask

```
import pandas as pd
from tabulate import tabulate
df = datasets["Car_sales"].copy()
print(tabulate(df[(df["Fuel_capacity"] > 15) & (df["Sales_in_thousands"] > 20)], headers="keys", tablefmt="html", showindex=False))
```

Manufacturer	Model	Sales_in_thousands	__year_resale_value	Vehicle_type	Price_in_thousands	Engine_size	Horsepower	Wheelbase	Width	Length	Curb_weight	Fuel_capacity	FuelL
Acura	TL	39.384	19.875	Passenger	28.4	3.2	225	108.1	70.3	192.9	3.517	17.2	
Audi	A4	20.397	22.255	Passenger	23.99	1.8	150	102.6	68.2	178	2.998	16.4	
Buick	Century	91.561	12.475	Passenger	21.975	3.1	175	109	72.7	194.6	3.368	17.5	
Buick	Regal	39.35	13.74	Passenger	25.3	3.8	240	109	72.7	196.2	3.543	17.5	
Buick	Park Avenue	27.851	20.19	Passenger	31.965	3.8	205	113.8	74.7	206.8	3.778	18.5	
Buick	LeSabre	83.257	13.36	Passenger	27.885	3.8	205	112.2	73.5	200	3.591	17.5	
Cadillac	DeVille	63.729	22.525	Passenger	39.895	4.6	275	115.3	74.5	207.2	3.978	18.5	
Chevrolet	Lumina	24.629	10.31	Passenger	18.89	3.1	175	107.5	72.5	200.9	3.33	16.6	
Chevrolet	Monte Carlo	42.593	11.525	Passenger	19.39	3.4	180	110.5	72.7	197.9	3.34	17	
Chevrolet	Camaro	26.402	13.025	Passenger	24.34	3.8	200	101.1	74.1	193.2	3.5	16.8	
Chevrolet	Impala	107.995	nan	Passenger	18.89	3.4	180	110.5	73	200	3.389	17	
Chrysler	Sebring Conv.	32.775	14.18	Passenger	24.495	2.5	168	106	69.2	193	3.332	16	
Chrysler	Concorde	31.148	13.725	Passenger	22.245	2.7	200	113	74.4	209.1	3.452	17	
Chrysler	Cirrus	32.306	12.64	Passenger	16.48	2	132	108	71	186	2.911	16	
Chrysler	300M	30.696	nan	Passenger	29.185	3.5	253	113	74.4	197.8	3.567	17	
Dodge	Stratus	71.186	10.185	Passenger	20.23	2.5	168	108	71	186	3.058	16	
Dodge	Intrepid	88.028	12.275	Passenger	22.505	2.7	202	113	74.7	203.7	3.489		

Fig.9: CSV Demonstration

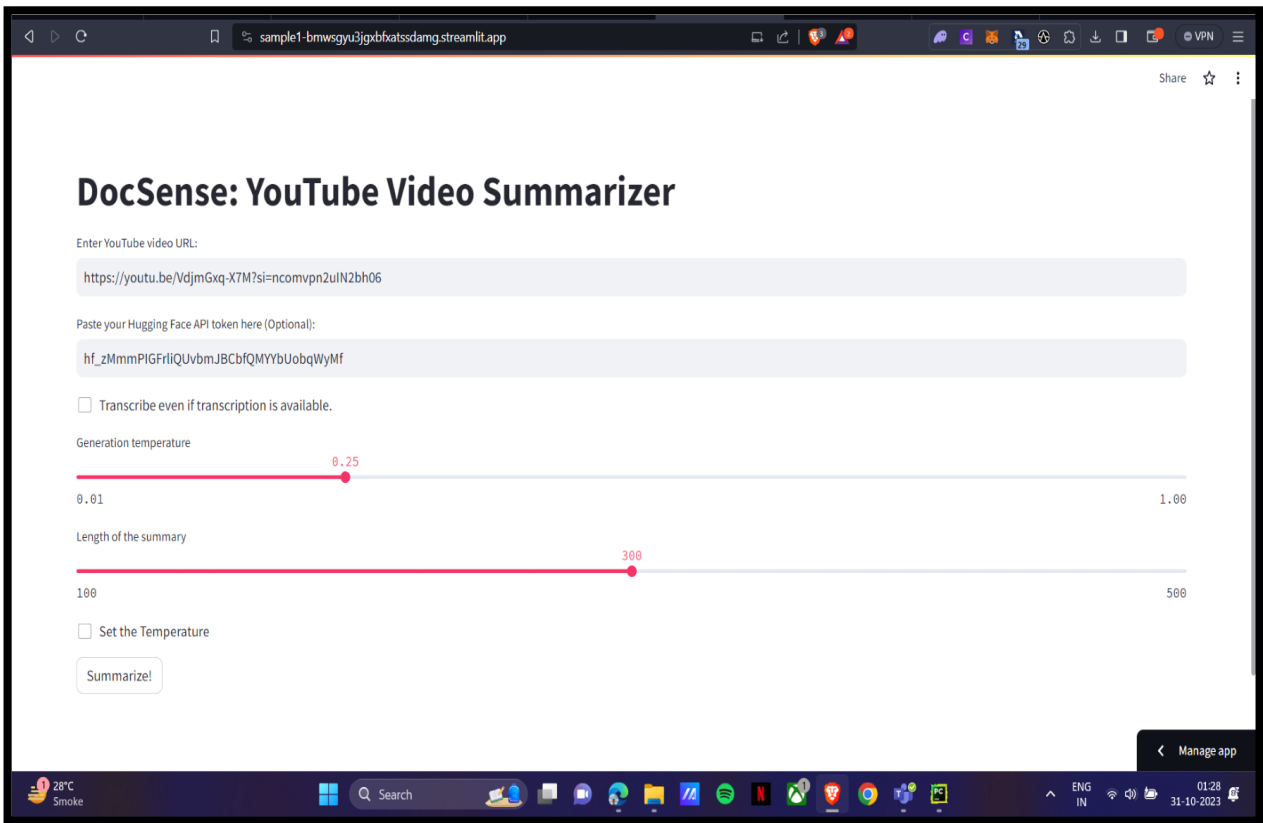


Fig.10: YouTube Summarizer Demonstration

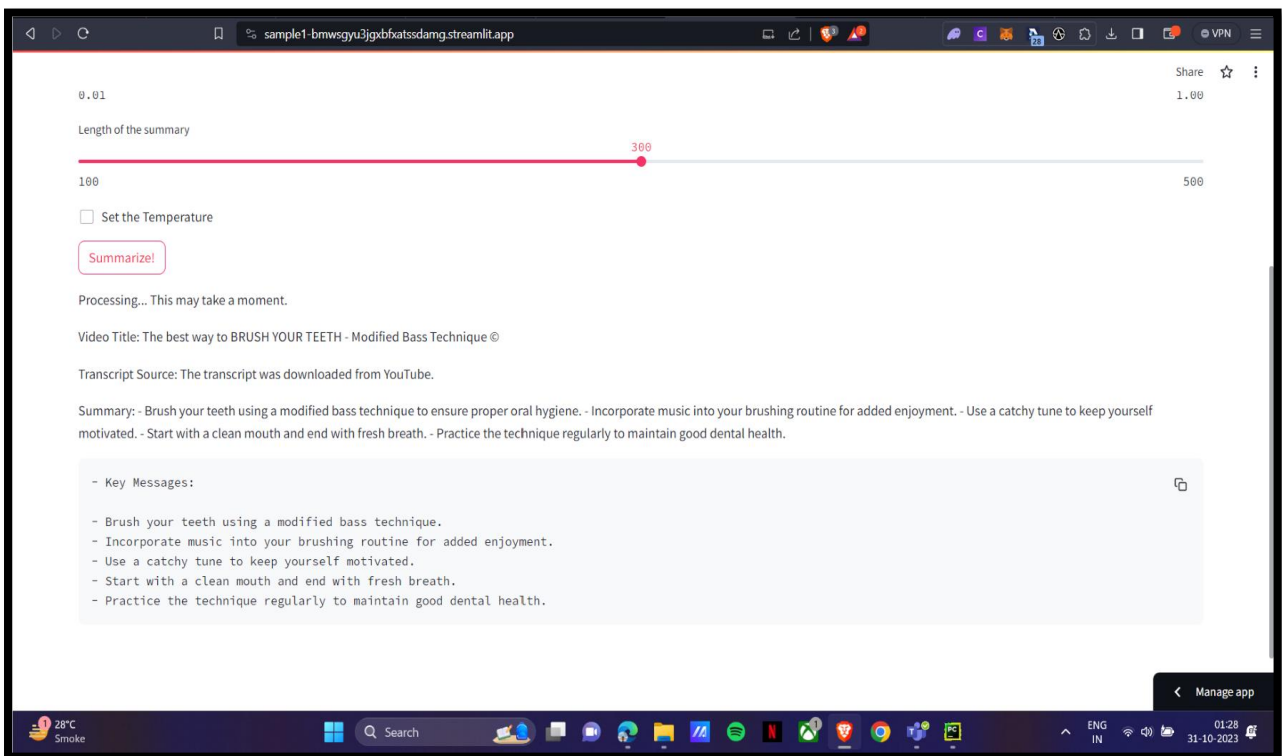


Fig.11: YouTube Summarizer Demonstration

5. CONCLUSION AND FUTURE SCOPE

DocSense stands out as a ground-breaking technology that bridges the divide between conventional document analysis and the constantly growing volume of digital data. Its benefits provide an indication of the revolutionary potential of AI in information extraction. The path does not end here, even while the current limits highlight the necessity for careful deliberation in its application. The future roadmap calls for ongoing improvement with the goal of fully using AI's potential to assist researchers, decision-makers, content producers, and knowledge disseminators in a variety of fields. With an eye towards the future, DocSense pledges to change and grow to satisfy the changing demands of a society that is becoming more and more focused on data. DocSense will continue to innovate, adapt, and improve user experience along the way. With its ability to accommodate many content kinds, languages, and communication idioms, it is positioned to influence how data analysis, information sharing, and dynamic interactions are conducted in the digital era. DocSense is headed towards becoming an essential tool for professionals in a variety of fields, thanks to its persistent dedication to satisfying the changing demands of both users and the larger data-driven world.

References

- [1] Partha Pratim Ray, "ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope", *Internet of Things and Cyber-Physical Systems*, Volume 3, Pages 121-154, 2023, 2023.
- [2] Adamopoulou, E., Moussiades, L., "An Overview of Chatbot Technology. In: Maglogiannis, I., Iliadis, L., Pimenidis, E. (eds) *Artificial Intelligence Applications and Innovations*", IFIP Advances in Information and Communication Technology, vol 584. Springer, 2020.
- [3] A. M. Rahman, A. A. Mamun and A. Islam, "Programming challenges of chatbot: Current and future prospective," *IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, Dhaka, Bangladesh, pp. 75-78, 2017.
- [4] B. Islam, M. Iqbal, G. Ubakanma and S. van der Vliet-Firth, "Skillbot: A Conversational Chatbot based Data Mining and Sentiment Analysis", *Human-Centered Cognitive Systems (HCCS)*, Shanghai, China, pp. 1-10., 2022.
- [5] Touvron, Hugo, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov et al. "Llama 2: Open foundation and fine-tuned chat models." *arXiv*, Cornell university, 2023.
- [6] Roumeliotis, Konstantinos I., Nikolaos D. Tselikas, and Dimitrios K. Nasiopoulos. "Llama 2: Early Adopters' Utilization of Meta's New Open-Source Pretrained Model." *arXiv*, Cornell university, 2023.
- [7] Awasthi, K. Gupta, P. S. Bhogal, S. S. Anand and P. K. Soni, "Natural Language Processing (NLP) based Text Summarization - A Survey," *6th International Conference on Inventive Computation Technologies (ICICT)*, Coimbatore, India, pp. 1310-1317, 2022.
- [8] Mandana Vaziri, Louis Mandel, Avraham Shinnar, Jérôme Siméon, and Martin Hirzel, "Generating chat bots from web API specifications", *Proceedings of the ACM SIGPLAN International Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software Association for Computing Machinery*, New York, NY, USA, 44-57, 2017.