

Intelli-Doc Navigator: Revolutionizing Document Interaction for the Visually Impaired

Abhay Chopde¹, Siddhant Tambatkar², Abaan Tamboli³ and Sanika Thorbole⁴

¹⁻⁴Vishwakarma Institute of Technology, Pune

Email: abhay.chopde@vit.edu, siddhant.tambatkar21@vit.edu, abaan.tamboli211@vit.edu, sanika.thorbole21@vit.edu

Abstract—The Intelli-Doc Navigator presents a revolutionary approach in document interaction, by seamlessly integrating advanced NLP (natural language processing) techniques with voice command control. By harnessing the ability of new robust technologies like Lang-Chain, semantic similarity matching, and dynamic response generation, this system provides users with an interactive platform for engaging with PDF documents. Its new methodology ensures precise retrieval and analysis of textual content which enhances the efficiency and accuracy in document navigation. The proposed system includes natural language interaction, multi-document support, and voice command control which makes it a great platform in various sectors like academics, research, and professional domains. It revolutionizes document navigation, allowing users to extract insights and traverse dense textual content with unprecedented efficiency and sophistication. This system's emphasis on seamless integration, innovative techniques, and transformative powers highlights its potential to reinvent accessibility and productivity in document interaction.

Index Terms— Lang-Chain, Artificial Intelligence, LLM, NLP, Voice-Control.

I. INTRODUCTION

In the digital era, the document navigation becomes more and more common in many areas. The old-style mode often becomes the bottleneck, for example, it loses 2. 2 billion people in the world whose eyesight are not normal as per WHO and similar numbers of people are given by ResearchGate journal article showed in fig 1. IntelliDoc Navigator has devised an entirely new method of interacting with documents with different formats using of language and voice commands. High features such as text to speech conversion, document summary, semantic analysis and customisable interfaces make it easy and simple for the visually impaired users. Moreover, the NLP (natural language processing) techniques and voice command control that are applied in the current time completely change the way visually impaired persons manage textual elements and provide a way for seamless and natural interaction with them, retrieval and utilization of different kinds of documents.

Through the use of text-to-speech conversion, intelligible speech is generated, producing a natural and coherent rendering of information in documents by informing the patient auditorily from the textual information. Text summarization engines are designed to give a cohesive view of lengthy texts and allow for a straightforward understanding by the visually impaired. Also, incorporation of semantic analysis and tailor-made interfaces increases the systems capacity of understanding the sense of the text correctly and providing the content of different needs and tastes. Besides accessibility the Intelli-Doc Navigator is also instrumental in educational organizations, research projects, business and other activities. In educational contexts, with the help of this technology the learning environments can become more inclusive by providing resources for visually impaired

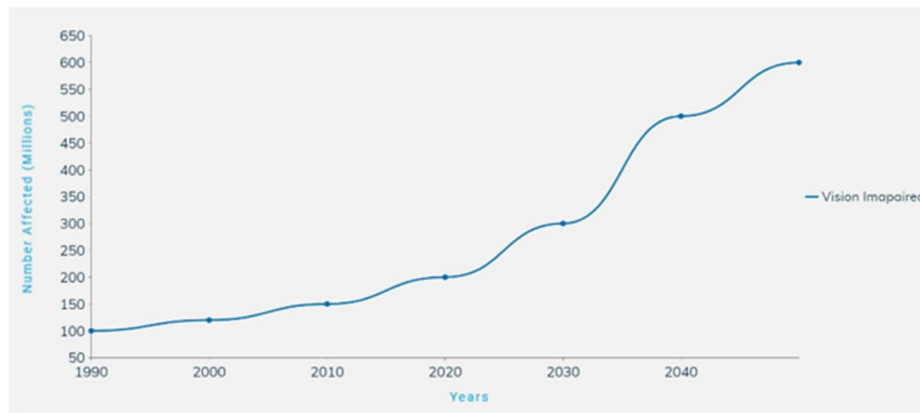


Figure 1 Global trends and predictions of numbers of people who are moderately and severely vision impaired, from 1990 to 2050

students which they may not otherwise have access to. Research, it assists in fast information acquisition and information process, therefore increasing the speed of the work of the people with visual handicaps. It improves both the document management processes and allows visually impaired professionals to work more effectively with textual content that they cannot read, and as a result, speeds up the process of making decisions and smooths the workflow.

The Intelli-Doc Navigator's advanced features extend beyond accessibility enhancements for visually impaired users. Additional functionalities such as multi-modal interaction, real-time collaboration, and integration with accessibility standards further enhance usability and inclusivity. Multi-modal interaction enables users to interact with the system using voice commands, touch gestures, or keyboard shortcuts, providing flexibility and accommodating different user preferences. Real-time collaboration features allow users to collaborate on documents with others and provide feedback, promoting inclusivity and fostering collaboration among users with diverse abilities. Integration with accessibility standards such as the Web Content Accessibility Guidelines (WCAG) ensures that the Intelli-Doc Navigator meets the needs of visually impaired users and adheres to best practices in accessibility design and development. Navigator not only empowers visually impaired users to navigate and interact with textual content more effectively but also finds applications across various domains, contributing to a more inclusive and accessible digital ecosystem.

II. LITERATURE SURVEY

This research examines how LLMs might be used to quickly construct applications, with a particular emphasis on the open-source software library LangChain. Among the document's Its fundamental elements, such as its chains and parts, serve as interchangeable pipelines and modular abstractions[1]. The research looks at the fundamental components of LangChain and shows how it might help accelerate the creation of LLM-based applications. This package allows voice assistants to function in offline mode. The Speech Recognition library in Python uses APIs such as CMU Sphinx, Google Cloud Speech API, and Snowboy Hotword Detection to translate voice to text. The four stages of a voice assistant's operation are text-to-speech output, natural language processing (NLP), automated speech recognition, and necessary business logic via hooks[2]. Additionally, a comprehensive assessment of the literature is available in the research area of voice assistants and smart personal assistants, with a focus on PICOC (population requirements, action, contrast, outcomes, and meaning). The assessment examined more than 1250 scientific publications that were released in the previous six years, highlighting 131 of the most significant works and outlining unresolved issues and difficulties. The study discusses the use, styles, structures, security and privacy issues, and present state of voice assistants[3]. Adopting Biolchini's principles for structural mapping investigations resulted in a hierarchical review process. An open-source architecture that may be scaled to any industry or organization is presented in a different paper. Its goal is to satisfy customer demands for instant service and question resolution. For demonstration purposes, the chatbot can respond to user inquiries by retrieving up-to-date information from the notice board of Birla Vishvakarma Mahavidyalaya (BVM) Engineering College. The model makes use of Google's Flan T5 XXL language model, which preserves the context of prior communications and makes use of it to forecast responses to inquiries in the future[4]. Information search and retrieval from PDF documents has become more difficult due to the explosive expansion of digital products. Using Streamlit, a web application framework, users can interact with PDFs and

search and retrieve documents more easily, revolutionizing Natural Language Processing (NLP). With Streamlit, you can make unique, eye-catching online apps without having to know how to use conventional web development frameworks like HTML and CSS. The application's back-end is made up of Open AI Embeddings and Large Language Models according to the application architecture. Streamlit manages the front-end portion of the online application where text inputs and user messages are received, contributing to the creation of an engaging and attractive user interface[5]. Additionally, the study's systematic review examines how well Large Language Models (LLMs) use ChatGPT and Lang Chain to solve mathematical word problems and numerical computational issues. The goal of the study is to analyze how well these LLMs function by looking at various question types and error types in the context of arithmetic problems. Prompt analysis, the process of analyzing and comprehending a given prompt or question in multiple contexts—often in the context of natural language processing (NLP) tasks or AI models—is the primary methodology employed in this work. The work uses many mathematical questions to provide an insightful examination of ChatGPT and LangChain[6]. Apart from investigating the possibilities of LLMs across many fields, scientists have created an intuitive framework called LLM-Adapter to make it easier to apply Penetration and Extraction (PEFT) techniques in research and real-world settings. For their experiments, they employ LLaMA, BLOOMz, and GPT-J as basis models. As representative candidates, they choose Prefix-Tuning, Series Adapter, LoRA, and Parallel adapter. For fine-tuning tests, a batch size of 16 is kept constant for consistency. With the use of the fine-tuning datasets, each PEFT method is optimized over three epochs. For the math or commonsense reasoning tasks, the authors optimize a single model and assess its performance across all relevant datasets[7]. Together, these examples demonstrate the wide range of uses and continuous developments in artificial intelligence and natural language processing, from academic research and customer service automation to application development. The significance of inventive frameworks, methodical literature studies, and comprehensive assessment techniques in advancing the skills and comprehension of LLMs and associated technology is emphasized. [8] One of the work discussed the impact of large language models like chatgpt on raising concerns about reliability, academic integrity and potential misuse in exams. It says that as chatgpt generates convincing texts still it falls short for meeting standards expected in exams of certified nuclear medicine physicians.[9]The other study delves into the intersection of LLM with security and privacy, showcasing their positive impact on code and data privacy. They categorized the findings into ‘The Good’, ‘The Bad’ and ‘The Ugly’ highlighting the need for further research.[10] Study explores the need of accurate detection of LLM generated text and crucial for bolstering trust in NLG systems. It delves into data acquisition strategies, feature selection techniques and statistical disparities.[11]Paper talks about the successful synthesis of 400g of the novel energetic heterocycle LLM-105, boosting 81% of energy of HMX and thermal stability.[12]There are some studies which investigates the biological functions of Arabidopsis LLM-domain B-GATA transcription factors, focusing on GNC,GNL,GATA15,GATA16,and GATA17L. By mutant analysis they revealed their roles in greening , floral organ initiation and senescence. [13] Using a solvent-slurry process with EPDM as a binder a new insensitive booster explosive, LLM-105/EPDM , based on 2,6-diamino-3,5-dinitropyrazing-1-oxide was developed. Also its shock wave sensitivity was lower than PBXN-5 and PBXN-7. [14] one of the study explores the emergence and application of chatbots as learning and assistance tools. It delved into design and development platforms and applications of it. Also discussed the history of Chatbots.[15] The other survey explores recent advancements in chatbots, focused on their implementation while utilizing Artificial Intelligence and Natural Language Processing. It addressed challenges such as contextual understanding and gender biases.

III. METHODOLOGY

The methodology for querying information from PDF files or books integrates advanced technologies to enhance accessibility and usability for visually impaired individuals, as depicted in Figures 2, and 3. Initially, upon user interaction, the system employs sophisticated techniques such as Optical Character Recognition (OCR) or specialized libraries like PyPDF to extract textual content from uploaded documents, as illustrated in Figure 3. This extracted data then undergoes preprocessing to ensure compatibility with subsequent stages of analysis. Data chunking is implemented to divide the extracted text into smaller, more manageable segments, typically containing around 2000 characters each. This chunking process, depicted in Figure 3, is crucial for overcoming input limitations imposed by subsequent models, allowing for efficient handling and analysis of the textual data. Subsequently, embeddings are generated for each text segment to capture semantic information, forming a repository in a knowledge base or vector store for rapid retrieval and comparison during query processing.

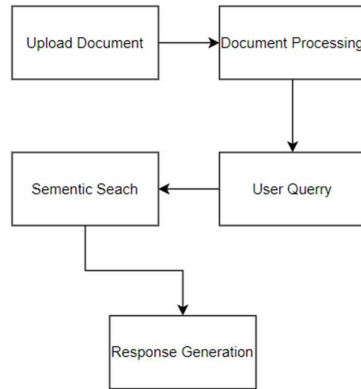


Figure 2 Intelli-Doc's Brief Overview

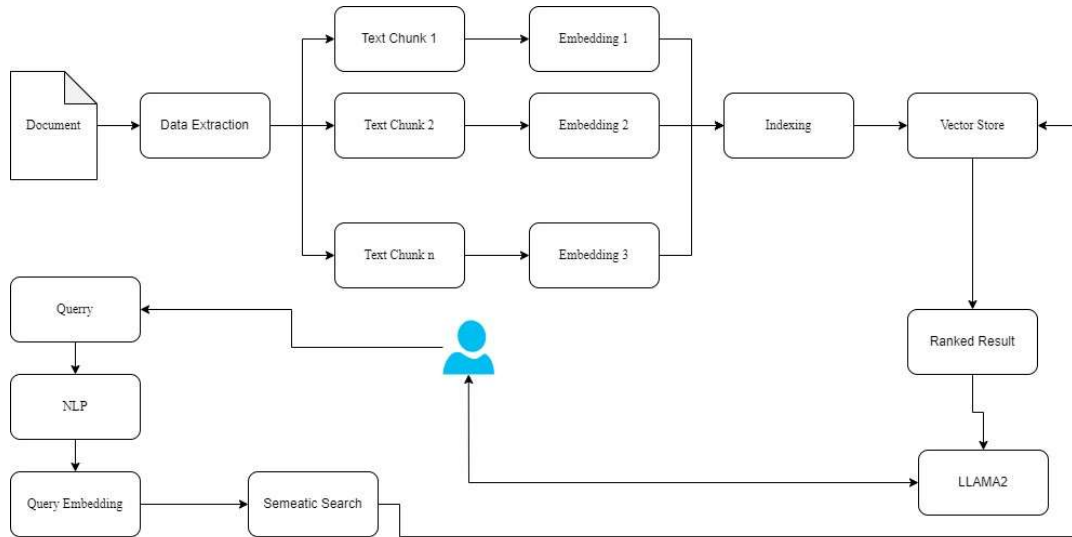


Figure 3 Intelli-Doc's Internal Working

Once the data preprocessing and embedding generation stages are complete, the system becomes capable of responding to user queries effectively. When a user submits a query or asks a question, the system initiates a semantic search process within the knowledge base. This involves comparing the embeddings of the user's query with the embeddings stored in the knowledge base to identify text segments exhibiting semantic similarity to the query. Through this semantic search mechanism, the system retrieves relevant information that aligns with the user's query. Figure 3 shows how the top-ranked results, based on similarity scores or query relevancy, are picked and presented to the user.

To improve the user experience and provide contextually relevant replies, the top results together with the user's query prompt are given as an input to a natural language processing system. Here, the LLAMA-2 language model, the newest one created by Meta AI, is demonstrated in Table 1. LLAMA2 provides a rich set of features for research and commercial applications and can be equipped with the number of pre-trained models that are optimized for different circumstances and computational resources. The fine-tuned models, which are trained to be highly proficient in dialog completion and question answering tasks, are deployed to deliver precise and contextual responses in reply to user inquiries. The decision to rely on the fine-tuned models is driven by the specific real-life scenarios, like question answering from PDFs or books for example. Table 2 provides the performance comparison of LLAMA2 with other benchmark models such as open-source and a few closed-source models which demonstrates that it has high effectiveness and versatility and can be used for most NLP applications. Furthermore, advanced NLP models like LLAMA2 are among the techniques that are being

integrated together with the use of Google API for speech-to-text and text-to-speech capabilities. Google APIs are employed to transform voice input from users into textual requests (speech-to-text) and to convert the textual answers from the system into a voice output for users (text-to-speech). Integration of the system raises the level of accessibility by making the visually challenged users interact with the system by voice commands and receive the auditory feedback while operating, thus enabling the blind people to have a great interaction with documents.

TABLE I. PARAMETER STUDY

Model Size (Parameters)	Pretrained	Fined-Tuned for chat use cases
7B	Model architecture: Pre training Tokens: 2 trillion Context Length: 4096	Data collection for helpfulness and safety: Supervised fine-tuning: over 1,00,000 Human preferences: Over 10,00,000
13B		
70B		

TABLE II. MODEL COMPARISON

Benchmark (Higher is Better)	MPT (7B)	Falcon (7B)	Llama2 (7B)	Llama2 (13B)	MPT (30B)	Falcon (40B)	Llama-1 (65B)	LLAMA-2 (70B)
MMLU	26.8	26.2	45.3	54.8	46.9	55.4	63.4	68.9
TriviaQA	59.6	56.8	68.9	77.2	71.3	78.6	84.5	85.0
Natural Questions	17.8	18.1	22.7	28.0	23.0	29.5	31.0	33.0
GSM8K	6.8	6.8	14.6	28.7	15.2	19.6	50.9	56.8
HumanEval	18.3	N/A	12.8	18.3	25.0	N/A	23.7	29.9
AGIEval	23.5	21.2	29.3	39.1	33.8	37.0	47.6	54.2
BoolQ	75.0	65.8	77.4	81.7	79.0	83.1	85.3	85.0
HellaSwag	76.4	74.1	77.2	80.7	79.9	83.6	84.2	85.3

Furthermore, the methodology includes the integration of other state-of-the-art features, like multi-modal interaction and real-time collaboration. The multi-modal interaction provides users with the ability to use voice commands, touch gestures, or keyboard shortcuts for system interaction to provide flexibility and accommodate with different user preferences. The capability of real-time collaboration in such apps enables users to involve other people in working on a single document together, provide feedback and facilitate collaborating interactions among people with different abilities. These properties are smoothly integrated into the system to afford improved user experience and pave the way for efficient document handling for visually impaired persons.

IV. RESULT AND DISCUSSION

The integration of LangChain as the primary component present in the architecture of the proposed system has been vital in terms of handling and retrieving textual information from PDF files or books. LangChain's advanced language processing capabilities were an indispensable factor that ensured that the system became more effective and accurate after extensive experimentation and observations. An important finding from the implementation is the enhanced semantic meaning and context awareness that was attained through LangChain's language processing capabilities. The different test settings demonstrated that LangChain had a better grasp of the textual content and hence gave more accurate and contextually adhering responses to the users. Such improvement in semantic understanding has been the cornerstone of the way the user experience and satisfaction have risen.

Moreover, LangChain's adaptability and flexibility allowed for seamless integration within the architecture, facilitating smooth data flow and interaction with external components. This observation was particularly evident during the testing phase, where LangChain seamlessly integrated with other modules to streamline document

preprocessing, semantic search, and response presentation processes. The modular design and flexible APIs of LangChain contributed to a reduction in integration overhead and improved system performance. Furthermore, the integration of LangChain has led to a more robust and scalable architecture for processing and retrieving textual information. By leveraging LangChain's sophisticated algorithms and models, the system has become better equipped to handle diverse document types and content, resulting in improved retrieval accuracy and efficiency across different use cases.

Overall, the observations from the implementation highlight the significant impact of LangChain on system performance and functionality. Its advanced language processing capabilities, coupled with seamless integration within the architecture, have elevated the system's effectiveness in processing and retrieving textual information. Moving forward, further exploration and optimization of LangChain's features hold the potential to enhance the system's capabilities and enable more sophisticated applications in information retrieval and knowledge discovery.

V. CONCLUSION

The Intelli-Doc Navigator revolutionizes document interaction through its integration of advanced NLP and voice command control technologies. It enhances accessibility and productivity for users across various domains, particularly those with visual impairments or mobility limitations. The system's innovative features empower users to navigate complex textual content efficiently and seamlessly. Moving forward, there are opportunities for further development and optimization to enhance user experience and expand applications in diverse domains. Ultimately, the IntelliDoc Navigator represents a significant advancement in document interaction technology, promising a more inclusive and efficient approach to document navigation.

REFERENCES

- [1] Topsakal, Oguzhan, and Tahir Cetin Akinci. "Creating large language model applications utilizing langchain: A primer on developing llm apps fast." In International Conference on Applied Engineering and Natural Sciences, vol. 1, no. 1, pp. 1050-1056. 2023.
- [2] Gonge, Sudhanshu, Atishay Jain, Rahul Joshi, Deepali Vora, and Ketan Kotecha. "Voice Recognition System for Desktop Assistant." In Computational Vision and Bio-Inspired Computing: Proceedings of ICCVBIC 2022, pp. 675-690. Singapore: Springer Nature Singapore, 2023.
- [3] Darda, Pooja, and R. M. Chitnis. "voice assistant: a systematic literature review". an international bilingual peer reviewed refereed research journal. 2020.
- [4] Pandya, Keivalya, and Mehfuza Holia. "Automating Customer Service using LangChain: Building custom open-source GPT Chatbot for organizations." arXiv preprint arXiv:2310.05421 (2023).
- [5] Adith sreeram A S, Pappuri Jithendra Sai, 2023, an effective query system using llms and langchain, international journal of engineering research & technology (ijert) volume 12, issue 06 (june 2023).
- [6] Soygazi, Fatih, and Damla Oguz. "An Analysis of Large Language Models and LangChain in Mathematics Education." In Proceedings of the 2023 7th International Conference on Advances in Artificial Intelligence, pp. 92-97. 2023.
- [7] Hu, Zhiqiang, Yihuai Lan, Lei Wang, Wanyu Xu, Ee-Peng Lim, Roy Ka-Wei Lee, Lidong Bing, and Soujanya Poria. "LLM-Adapters: An Adapter Family for Parameter-Efficient Fine-Tuning of Large Language Models." arXiv preprint arXiv:2304.01933 (2023).
- [8] Alberts, Ian L., Lorenzo Mercolli, Thomas Pyka, George Prenosil, Kuangyu Shi, Axel Rominger, and Ali Afshar-Oromieh. "Large language models (LLM) and ChatGPT: what will the impact on nuclear medicine be?." European journal of nuclear medicine and molecular imaging 50, no. 6 (2023): 1549-1552.
- [9] Yao, Yifan, Jinhao Duan, Kaidi Xu, Yuanfang Cai, Eric Sun, and Yue Zhang. "A survey on large language model (llm) security and privacy: The good, the bad, and the ugly." arXiv preprint arXiv:2312.02003 1 (2023).
- [10] Tang, Ruixiang, Yu-Neng Chuang, and Xia Hu. "The science of detecting llm-generated texts." arXiv preprint arXiv:2303.07205 (2023).
- [11] Pagoria, Philip F. Synthesis, scale-up, and characterization of 2, 6-diamino-3, 5-dinitropyrazine-1-oxide (LLM-105). No. UCRL-JC-130518; CONF-9806105-. Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States), 1998.
- [12] Ranfil, Quirin L., Emmanouil Bastakis, Carina Klermund, and Claus Schwechheimer. "LLM-domain containing B-GATA factors control different aspects of cytokinin-regulated development in Arabidopsis thaliana." Plant physiology 170, no. 4 (2016): 2295-2311.
- [13] Xu, Wenzheng, Chongwei An, Jingyu Wang, Jun Dong, and Xiaoheng Geng. "Preparation and properties of an insensitive booster explosive based on LLM-105." Propellants, Explosives, Pyrotechnics 38, no. 1 (2013): 136-141.
- [14] Tamrakar, Rohit, and Niraj Wani. "Design and development of CHATBOT: A review." ResearchGate, Apr (2021).
- [15] Caldarini, Guendalina, Sardar Jaf, and Kenneth McGarry. "A literature survey of recent advances in chatbots." Information 13, no. 1 (2022): 41.