INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 09 ISSUE: 01 | JAN - 2025 SJIF RATING: 8.448 ISSN: 2582-3930

MitraBot : Your One Step Guide towards all Loans and Schemes

Sathvik N B Math
School of Computer Science and Engineering
Presidency University
Bengaluru, India
sathviknbmath@gmail.com

Adithya H Hegde
School of Computer Science and Engineering
Presidency University
Bengaluru, India
adithyahegde132@gmail.com

Vijayeendra N
School of Computer Science and Engineering
Presidency University
Bengaluru, India
swamyvijayeendra@gmail.com

V.Adithya
School of Computer Science and Engineering
Presidency University
Bengaluru, India
adith348@gmail.com

Mr. Pakruddin B
School of Computer Science and Engineering
Presidency University
Bengaluru, India
fakrubasha@gmail.com

***Abstract*— Simply, the digital governance accelerates citizen-government service interaction more with simple tools. In the paper, a chatbot will be developed using web scraping techniques dynamically collecting and updating needed information from government websites based on eGovernance. This will facilitate meaningful, intuitive conversations between users and the language model, Llama 3.2, making the access to public data and services even more interesting. All this will be collected in real time by the web scraper: knowledge base of the chatbot, including policies issued, deadlines covered by the government, citizen services offered, and many more. Inserting language comprehension and conversational capabilities of Llama 3.2 with gathering data in real time would be expected to create ease of interaction for the user, with minimum delay in information, and a realistic answer to citizens' inquiries. Discussions also include issues when web scraping on public websites is applied, legal perspectives, and efficiency in the integration of large language models into eGovernance systems. Results are readily available, responsive, and satisfactory concerning user satisfaction for AI-powered government service chatbots.**

Index Terms-Chatbot, eGovernance, Llama 3.2, Natural Language Processing

I. INTRODUCTION

E-governance is integration involving the transformation of the last few decades of digital technologies that transform through which governments view their relations with citizens. Nevertheless, information remains not so accessible, especially to multilingual regions and more so to not-so-technologically savvy citizens. This, therefore, makes pertinent the design of an approachable talking interface across this divide.

The purpose of a CHATBOT is to help answer user queries. CHATBOT is a computer program that processes a user's natural-language input and generates relatively smart, affluent, and intelligent responses sent back to the user.[1] Using the chatbot concept, users can easily interact with a government system. They are much more live, interactive, and intuitive compared to some search engine or static web portal where, based on a specific keyword input, users may receive their real-time, personalized responses. This situation introduces new hope with big language models like Llama 3.1, which possesses far greater capabilities than the usual state-of-the-art NLP, especially being much more contextually aware of understanding and producing human-like responses. This conceives a development for an e-governance-based chatbot on Llama 3.1 as a core NLP engine. The chatbot uses real-time data acquired from government websites and databases in the form of PDFs using LangChain. LangChain is a solution which helps us in the querying process and extracting information from PDFs. With its advanced NLP algorithms, it helps users to interact with the PDFs and makes the document search and retrieval very easy.[2] Meaning, citizens get prompt and reliable information instead of struggling with procedural bureaucratic forms.

II. LITERATURE REVIEW

Title - 01: Early Adopters' Utilization of Meta's New Open-Source Pretrained Model

Detail: Llama 3, Meta's latest open-source model, has seen early adoption across industries for applications like natural language processing, AI research, and software development. Its customizable architecture and powerful language generation capabilities have enabled developers to fine-tune it for specific needs, boosting productivity and innovation. [10]

Drawbacks: Some users report high computational demands and difficulties with fine-tuning, especially when deploying Llama 3 on limited hardware. Additionally, its open-source nature raises security and ethical concerns as misuse potential increases without proprietary controls.

Title - 02: Building Customized Chatbots for Document Summarization and Question Answering using Large Language Models with OpenAI, LangChain, and Streamlit

Detail: Using OpenAI's language models integrated with LangChain and Streamlit enables developers to create chatbots

tailored for document summarization and precise question answering. This framework supports customization, making it efficient for automating information retrieval and delivering concise summaries. [11]

Drawbacks: High computational costs and latency issues can hinder performance, particularly for large documents. Additionally, managing data privacy and ensuring accurate responses remain challenging when handling sensitive information.

Title - 03: GROQ ROCKS Neural Networks

Detail: GROQ processors bring high-speed, low-latency performance to neural network processing, optimizing deep learning workloads for real-time applications. Their unique architecture enables efficient handling of complex models, reducing time-to-inference and boosting AI capabilities. [12]

Drawbacks: Despite the speed, GROQ's specialized hardware is costly and requires significant infrastructure support, limiting accessibility. Additionally, compatibility with existing software stacks can be challenging, necessitating customized development efforts.

Title - 04: RAG (Retrieval-Augmented Generation)

Detail: RAG combines retrieval-based and generative models to enhance response accuracy, allowing it to pull relevant information from a knowledge base and generate coherent, context-aware answers. This approach is highly effective for applications in customer support and knowledge-intensive tasks. [13]

Drawbacks: RAG models require extensive computational resources and rely on high-quality, updated data sources, which can be costly to maintain. Additionally, they may struggle with response consistency, especially when dealing with ambiguous or poorly indexed data.

III. METHODOLOGY

Large Language Models represent complex artificial intelligence capable of understanding, generating, and manipulating human languages. They are based on deep learning architectures that generally involve or are built around the Transformer model, which allows them to process substantial amounts of text in learning patterns, context, and even linguistic structures. LLMs go one step further in learning the meanings of individual words, and also relationships among words, sentences, and broader textual elements that will permit them to construct coherent, contextually relevant responses.

They work by embedding layers that break down text into smaller units-known as tokens-and turn them into vectors. These vectors feed into multiple layers within the model, which include self-attention mechanisms for weighting the importance of a token in relation to others for contextual captures. Fully capturing the nuanced relationships found within these layers enables refinements to be made by the deeper layers of the model.

They get pre-trained on general text data and then fine-tuned to do translation tasks, summarization, or conversation. Along the way, they learn grammar, facts, reasoning, and even subtleties like tone and intent. Large language models like Llama 3.1 have great understandability and the capability to generate human-like text, making them very effective for chatbots, content creation, and automatic question answering applications [3]. They could handle complex representational language tasks, including multi-lingual and domain-specific interactions, which are very suitable for use cases in eGovernance, customer services, and education.

A. Llama 3.1

The newest, large language model released by the company is from Meta, formerly Facebook. Llama 3.1 broadens on the already published versions of Llama because it extends capabilities to better understand and generate text that closely resembles the form and structure of human language across a broad range of subjects. Like other LLMs, it is Transformer architecture-based, but Llama 3.1 is optimized for efficiency, so it can perform quite perfectly even on fewer computational resources than some of the large models similar to GPT-4. Llama 3.1 was subjected to the strongest training with a large, multilingual and diversified source dataset [3]. Therefore, it exhibits varying skills of performance in respect to different languages and domains of specialty. It enables flawless work in complicated situations-for example, legal document processing, scientific research, e-governance, etc. This also comes along with fine-tuning enhancements which have made it easier to adapt with minimal data for special tasks. Llama 3.1 has adapted features on reasoning, summarization, and even sentiment analysis, making it capable of doing many more things than other massive language models.

B. Llama 3.1 vs LLMs

B.1. Llama 3.1 vs GPT-4

Some of the most advanced LLMs developed in this AI landscape so far include Llama 3.1 and GPT-4, designed with highly sophisticated deep penetration natural language processing tasks at their core. Its focus on optimal utilization of resources also allows it to do so much better than stronger demands made by computations and is also extremely versatile for deployment at a large scale, including chatbots and real-time support in content creation provided by eGovernance services.

Comparatively, GPT-4, has been reported to possess wide scalability and generalization capabilities. Having been trained on a much bigger corpus than the previous models, GPT-4 is significantly more versatile in an enormously wider spectrum of tasks ranging from creative writing, reasoning, and coding.

That makes the strength of GPT-4 on its greater capacity for nuances in the responses and the depth it could exhibit while holding context in multi-turn conversations.[4]

Llama 3 405B performs approximately on par with GPT-4 (0125 API version). It outperforms GPT-4 in multiturn reasoning and coding tasks. However, it underperforms compared to GPT-4 in multilingual prompts (Hindi, Spanish, and Portuguese). [3]

B.2. Llama 3.1 vs Gemini 1.5

Llama 3.1 and Gemini AI are two releases of very high-functioning large language models with different explicit design advantages for some specific uses in AI. In contrast, Gemini AI by DeepMind is a multi-modal AI model that brings language processing together with a huge number of kinds of data - images, videos, and so on. So, Gemini AI may be used for solving impressively diverse tasks, moving far beyond the strict text-based interaction in order to solve challenging problems in such spheres as healthcare and autonomous systems to new types of content. Its multimodality makes it rather distinct from Llama 3.1, which is still much more text- and conversation-oriented. Of course, Llama 3.1 may perform exceptionally well with tasks of real-time natural language understanding and conversational dynamics, but one of the strengths of the Gemini AI technology is its ability to fuse different kinds of data toward ever more holistic and contextual problem solving.

While, Gemini 1.5 Pro achieves 100% recall up to 530k tokens and >99.7% recall up to 1M tokens [5]. Llama 3.1 models demonstrate perfect needle retrieval performance, successfully retrieving 100% of needles at all document depths and context lengths. We also measure performance on Multi-needle (Table 21), a variation of Needle-in-a-Haystack, where we insert four needles in the context and test if a model can retrieve two of them. Our Llama 3 models achieve near perfect retrieval results. [3]

B.3. Llama 3.1 vs Gemma 2

Among the newer state-of-the-art language models are Llama 3.1 and Gemma 2, built for various features of different use cases in the artificial intelligence landscape. The Llama 3.1 architectural design intends to produce context-aware responses in a wide scope of languages while carrying flexibility with fine-tuning for specific domains, making it well suited for very specialized tasks. While much larger in scope, Gemma AI will have much more complex data processing applications aside from its capabilities in natural language [6]. Llama 3.1, is basically an exercise in conversational ability, but Gemma AI would have far more robust capabilities with multitasking applications, handling complex analytics, and making it even stronger in applications such as enterprise business intelligence, predictive modelling, and many other things. While Llama 3.1 is fitted for effective and smart, intelligent natural conversations, the entire solution from Gemma AI integrates natural language processing with superior data interpretation to create an expertly crafted product-ideal for organizations looking for deep AI solutions.[6]

1.Data is stored in CSV files

The data will be stored inside the structured chatbot system through a structured CSV-based approach for efficient recall of data and thus reduction in dependability on the real-time web scraping [7]. The strategy reduces the need for constant, real-time scraping by external government websites that could be slow and sometimes unreliable because of changes in website structures or network instability.

For all the key fields like service names, procedures, eligibility criteria, documents needed, and links to official forms or guidelines, data of nearly all services through government, categorized and structured in each individual CSV file is given. In this way, it guarantees a stable and constant process for data fetching and hence enables the chatbot to respond in real-time while preventing delays usually incurred from querying via other external websites or APIs.[7]

Data in CSV refreshes constantly using either automated or semi-automated extraction of data, thus accurate and current without the problem of memory developing since constant web scraping goes on. This approach keeps structured data in such a fashion that it optimizes searchability on the fly by using user queries. With this system architecture is designed to locate the most relevant information in millicents leading to results in which the experience of the end user gets significantly improved.

2.Natural Language Processing

Advances in Natural Language Processing (NLP) have brought about a wide range of techniques and approaches that enable machines to comprehend and process human language. This section explores various methodologies that form the foundation of NLP and contribute to the transformation of machines' understanding of human language.

- We will employ the NL model Llama 3.1; this is based on best-in-class self-attention mechanisms and deep learning architecture as its underpinning. This subsequent section describes how it processes an incoming user input to break it down to extract information required from the CSV data storage system.

  A. Tokenization and Text Preprocessing:

  Tokenization serves as a fundamental preprocessing step in NLP. It involves breaking down a text into smaller units, typically words or sub words, referred to as tokens. Tokenization lays the groundwork for subsequent analysis, such as part-of-speech tagging and sentiment analysis [8].

  B. Part-of-Speech Tagging and Named Entity Recognition:

Part-of-speech tagging involves assigning grammatical categories (e.g., noun, verb, adjective) to words in a sentence. This technique aids in understanding the syntactic structure of a sentence. Named Entity Recognition (NER) identifies entities such as names of people, places, and organizations within text [8].

C. Syntax and Grammar Parsing:

Syntax and grammar parsing involve analyzing the grammatical structure of sentences. This includes identifying subjects, predicates, objects, and their relationships. Dependency parsing and constituency parsing are common techniques used for this purpose [8].

D. Question Answering and Information Retrieval

Question answering systems use NLP techniques to comprehend and answer questions posed in natural language. Information retrieval involves finding relevant documents or passages in response to a query [8].

E. Contextual comprehension and multi-turn dialogues: Multi-turn conversations are another new elaborated feature of Llama 3.1. After processing and answering the first question, follow-up questions from related items about the preceding interaction will be asked by the users. For instance, in relation to the question of application for a driving license, the user may ask, "Which documents are required?" In this aspect, Llama 3.1 is contextual about the conversation and recognizes that, "Does she still refer to the driving license application, and gives a relevant, extended answer about the required documents.". Answers from Llama 3.1 are, therefore in human interactive tones, so that whatever information is communicated will be accurate and readable.[9]

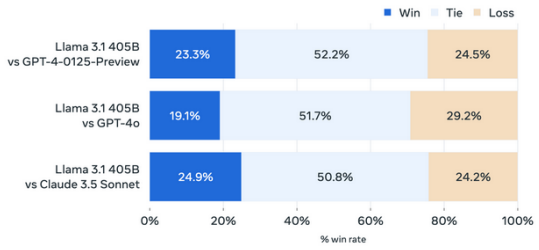

Fig. 1. Human evaluation of Llama 3.1

Model	MGSM	Multilingual MMLU
Llama 3 8B	**68.9**	**58.6**
Mistral 7B	29.9	46.8
Gemma 2 9B	53.2	–
Llama 3 70B	**86.9**	**78.2**
GPT-3.5 Turbo	51.4	58.8
Mixtral 8×22B	71.1	64.3
Llama 3 405B	**91.6**	83.2
GPT-4	85.9	80.2
GPT-4o	90.5	**85.5**
Claude 3.5 Sonnet	**91.6**	–

Fig. 2. Llama Multi Lingual test scores

Model	MGSM	Multilingual MMLU
Llama 3 8B	**68.9**	**58.6**
Mistral 7B	29.9	46.8
Gemma 2 9B	53.2	–
Llama 3 70B	**86.9**	**78.2**
GPT-3.5 Turbo	51.4	58.8
Mixtral 8×22B	71.1	64.3
Llama 3 405B	**91.6**	83.2
GPT-4	85.9	80.2
GPT-4o	90.5	**85.5**
Claude 3.5 Sonnet	**91.6**	–

Fig. 3. Proeficieny test scores of Llama 3.1

The eGovernance chatbot system has been built up through some methodology advances and is fundamentally dependent on Llama 3.1, being an improvement of its natural language processing capacities. This has enhanced information accuracy provided by the chatbot but in the process, has made it efficient and friendly to the users when considering an enormously large number of different user inputs, languages, and contexts. The next sections describe the main developments achieved with Llama 3.1 and the hybrid combination with a structured CSV-based data storage system.

- Deeper contextual comprehension

Llama 3.1 brings significant improvement in understanding context, which is critical for the proper and timely provisioning of information within the eGovernance context. Some of the core issues related to the development of a public service chatbot will be handling variations in and vagueness of user inputs. Users, for the most part, are inputting partial, vague, or ambiguous queries while searching for government services. For instance, it is possible for a claimant to ask how one obtains benefits without including either the actual form of benefit or the qualification requirements. That is where Llama 3.1's advanced architecture comes into play. The self-attention mechanism of Llama 3.1 chooses the essential linguistic elements and tends to interpret the larger context of any conversation very well. The model may also infer the intent of the user by merely selecting which information is likely to be sought, from some partial inputs. Llama 3.1, while analyzing

surrounding texts or through deep training on vast datasets, can fill gaps in user queries and create a response the user would be likely to need - even if the query is incomplete or contains less specificity. [14]

For example, if the user is vague in his query Llama 3.1 can continue to probe for clarification of what the user intends or at least what choices within a more general category to which may be referring-for example, different kinds of benefits available.

This version, Llama 3.1 is also multi-turn conversational enabling the chatbot to carry context across multiple turns. So, responses created by the system are able to build on previous queries retaining relevant information for the conversation, and so with a user experience. This retention of context is critical in eGovernance scenarios where users will have to ask several related questions-perhaps following up on required documents or eligibility after an initial question about an application process. [15]

- Data Retrieval Efficiency

The stored CSV file system improved the performance of the chatbot mainly by retrieving data and response generation. The systems designed using the technique of web scraping have problems like latency, modification of website structure, and sometimes at specific times sources down. This leads to degradation in the response time and reliability of the system. [16]. To address these limitations, the eGovernance chatbot saves all relevant information into a formatted CSV file.

This is a pre-indexed static repository of every significant information related to the services offered by governments. It is an accumulation of information regarding policy changes, application steps, document requirements, and many other details. Since everything is preserved at one place, data is fetched in real time, with lookups performed within the CSV file rather than fetching data from some other websites and databases in real time.

This structured approach enables the system to allow latency reduction, thereby delivering responses quickly, which would be important to the user who wants answers to questions right away. The data in CSV will then be refreshed periodically to ensure that the chatbot delivers information timely and appropriate without having to constantly conduct real-time scrapes. This does not only make the system stable but also assures a high scalability since the volume of inquiries handled is not degrading the performance of the system.

- Multi-Language Support

One of the key features for Llama 3.1 is robust multi-language support for e-Governance chatbots to interact with a heterogeneous population. Particularly, for countries having more than one official language and linguistic diversity, it becomes extremely important that government services must be usable by all citizens, irrespective of the preferred language. Llama 3.1 can understand and respond to users in several languages so the chatbot can be able to interact with users in their preferred native language and thus increase accessibility and inclusivity. Trained on immense multilingual datasets that enable the full usage of Llama 3.1 in conversation across multiple languages, the model presents vast efficiency not only in understanding and generating text for the major global languages but also exceling well when dealing with regional languages, hence being much more useful for such countries where the citizens require services in their respective local languages rather than a widely spoken national language or the English language. The chatbot automatically recognizes the language in which the query has been submitted and, therefore responds in the same language, thus offering a seamless experience to the user. This particular feature knocks down the language barriers and gives critical government information to a more substantial population of the user base. Second, Llama 3.1 has contextual knowledge, meaning it can work with hard and multi-turn conversations in multiple languages all while holding context over numerous interactions. [17]

IV. EXPERIMENTAL RESULTS

The results from our chatbot experiments show that it effectively answers user questions by pinpointing and retrieving relevant loans and schemes tailored to different sectors. When a query is received, the chatbot actively searches through a wide range of resources, including PDF documents, to pull out specific information related to the sector in question—be it agriculture, education, small businesses, or other focused areas. The responses from the chatbot are customized to match user intent, offering clear yet thorough details about available government schemes or loan options, eligibility requirements, and potential benefits in each sector. This ability to provide sector-specific responses emphasizes the chatbot's role as a valuable eGovernance tool, allowing users to quickly find relevant information without the need to navigate through complicated documentation. The findings highlight the system's accuracy, responsiveness, and effectiveness in improving user interaction and satisfaction.

Unnat Bha	Employme	Subsidy on	Project dur	Provides EPF subsidy to employers who create new jobs and provide social security benefits to employees.
Pradhan M	Agriculture	Subsidized	Ongoing	Supports farmers to manage the supply chain of TOP (Tomato, Onion, Potato) crops to stabilize prices.
Pradhan M	Agriculture	Grant/subs	Project dur	Financial support for the development of horticulture, including fruits, vegetables, and spices.
National Fa	Organic Fa	Subsidized	3-5 years	Promotes organic farming practices and provides assistance for certification and marketing.
National Sl	MSME/Ma	Subsidy-ba	Project dur	Provides employment and skill development through solar-powered charkha (spinning wheel) clusters.
Swabhima	Fisheries	Subsidized	5-10 years	Financial support for fisheries, fish farmers, and fishery infrastructure development.
PM SVANic	Agriculture	Subsidized	5 years	Promotes bamboo cultivation and supports bamboo-based industries for sustainable development.
National Cl	Innovation	Grant-base	Project dur	Provides seed funding for startups and innovators in India to scale their ideas and products.
Sampoorn	Social Welf	Subsidized	Project dur	Provides financial assistance for aids and appliances for disabled individuals to enhance mobility.
Rashtriya \	Healthcare	Subsidized	Ongoing	Aims to promote AYUSH (Ayurveda, Yoga, Unani, Siddha, and Homeopathy) healthcare systems across India.
PM POSHA	Infrastruct	Grant-base	Project dur	Develops infrastructure in border areas to support socio-economic growth and strengthen border security.
Stand Up I	Rural Deve	Grant-base	Ongoing	Aims to develop model villages with basic facilities, including health, education, and sanitation.
National A	Mining Sec	Grant-base	Project dur	Improves welfare in areas affected by mining, focusing on infrastructure, education, and healthcare.
Prime Min	Health Insu	Subsidized	1 year (ren	Provides free health insurance cover up to ?5 lakh per family per year for secondary and tertiary care.
Pradhan M	Education	Grant-base	Project dur	Supports international research collaborations and strengthens academic partnerships globally.
Venture Ca	Higher Edu	Grant-base	Project dur	Supports quality improvement in higher education institutions through funding for infrastructure and teaching.
Mudra Yoj	Women En	Free platfo	Ongoing	Online platform for women entrepreneurs to market their products, promoting economic empowerment.
Pradhan M	Environme	Project-ba	Project dur	Aims to rejuvenate and restore the river Ganga through infrastructure, waste treatment, and awareness.
Mission In	Rural Deve	Project-ba	Ongoing	Connects higher education institutions with villages to address rural problems with scientific solutions.
Atal Bhujal	Social Welf	Free/subsi	3 years	Provides LPG connections to BPL families to reduce indoor pollution and encourage cleaner cooking fuel.
Rashtriya (	Agriculture	Subsidy-ba	Ongoing	Aims to provide water to every farm by improving irrigation infrastructure and water use efficiency.
Swachh Bh	Skill Develo	Free trainii	1-3 years	Offers entrepreneurship education and training to foster self-employment among youth.
National Ri	Agriculture	Subsidy-ba	Seasonal	Supports increasing food grain production through various initiatives, including technology and research.
PM-CARES	Skill Develo	Subsidy-ba	Project dur	Aims to improve skill development infrastructure to meet the demand for a skilled workforce in India.
Pradhan M	Financial Ir	0% (no int	Ongoing	Focuses on bringing banking services to rural areas without bank branches, promoting financial inclusion.
Shyama Pr	Micro-ente	7% 1 year	(ren	Provides working capital loans to street vendors to help them revive their businesses post-COVID-19.
Pradhan M	Child Welf	Grant-base	Project dur	Aims to rehabilitate child laborers through education and vocational training programs.
National H	Rural Empl	Wage-base	100 days/y	Provides employment and food security in rural areas through wage employment for the unemployed poor.
Mahatma (	Social Welf	Free for eli	Project dur	Provides physical aids and assistive devices to senior citizens from BPL families.
Pradhan M	Education/	Free meals	School dur	Provides nutritious meals to school children to improve health and school attendance rates.
Ujjwala 2.(	Entreprene	8-10%	7 years	Provides loans to SC/ST and women entrepreneurs for setting up greenfield projects.
Pradhan M	Skill Develo	Grant-base	1-2 years	Provides financial support to industries to engage apprentices, improving workforce skills.
Pradhan M	MSME	Subsidized	3-5 years	Financial assistance to set up new micro-enterprises in rural and urban areas.
Jal Jeevan	Healthcare	Subsidized	Ongoing	Provides affordable generic medicines through Jan Aushadhi stores across India.
Integrated	Agriculture	Subsidized	1 cropping	Crop insurance scheme to protect farmers against crop loss due to natural calamities.
Mahila San	Agriculture	Grant-base	Project dur	Financial support for agribusiness projects, promoting entrepreneurship in agriculture.

Fig. 4. Sample Dataset

Fig. 5. Response from the Chatbot

V. CONCLUSION AND FUTURE WORK

Our system is a new way through which the user can obtain access to information about government loans and schemes in an eGovernance manner that changes the way. Since it processes language sophisticatedly, it parses and interprets PDF documents, which helps the system point out relevant information about the key aspects of every different inquiry, and hence the user receives the accurate, context-sensitive responses to meet his or her needs.

It will retrieve all the complex information that is going to decrease the complexity of searching lengthy documents in the government schemes and loans. Thus, through automation, it's easy for the users as they would not require any manual search from long lines of documents and information retrieval is very less hectic and time-consuming. It is also fairer because the chatbot presents information in a more understandable way; therefore, the application is helping people regardless of their level of inexperience in handling any form of government paperwork.

This brings about transparency and permits a timely and accurate distribution of information to users as it provides a foundation of an informed decision. As part of our eGovernance program, a chatbot connects the citizen with the public resources in more inclusiveness to gain access services from the government. Example of how technology allows better access to key information as it is faster and makes it democratic as the citizen gets closer to receiving service provision.

REFERENCES

[1] Tamrakar, Rohit & Wani, Niraj. (2021). Design and Development of CHATBOT: A Review.

[2] Sreeram a, Adith & Sai, Jithendra. (2023). An Effective Query System Using LLMs and LangChain. International Journal of Engineering and Technical Research. 12.

[3] Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., Mathur, A., Schelten, A., Yang, A., Fan, A. and Goyal, A., 2024. The llama 3 herd of models. arXiv preprint arXiv:2407.21783.

[4] Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S. and Avila, R., 2023. Gpt-4 technical report. arXiv preprint arXiv:2303.08774.

[5] Reid, M., Savinov, N., Teplyashin, D., Lepikhin, D., Lillicrap, T., Alayrac, J.B., Soricut, R., Lazaridou, A., Firat, O., Schrittwieser, J. and Antonoglou, I., 2024. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. arXiv preprint arXiv:2403.05530.

[6] Team, G., Riviere, M., Pathak, S., Sessa, P.G., Hardin, C., Bhupatiraju, S., Hussenot, L., Mesnard, T., Shahriari, B., Ramé, A. and Ferret, J., 2024. Gemma 2: Improving open language models at a practical size. arXiv preprint arXiv:2408.00118.

[7] C. Tapsai, "Information Processing and Retrieval from CSV File by Natural Language," 2018 IEEE 3rd International Conference on Communication and Information Systems (ICCIS), Singapore, 2018, pp. 212-216, doi: 10.1109/ICOMIS.2018.8644947.

[8] Rayhan, Abu & Kinzler, Robert & Rayhan, Rajan. (2023). NATURAL LANGUAGE PROCESSING: TRANSFORMING HOW MACHINES UNDERSTAND HUMAN LANGUAGE. 10.13140/RG.2.2.34900.99200.

[9] Sam, Kira & Vavekanand, Raja. (2024). Llama 3.1: An In-Depth Analysis of the Next Generation Large Language Model. 10.13140/RG.2.2.10628.74882.

[10] Roumeliotis, Konstantinos & Tselikas, Nikolaos & Nasiopoulos, Dimitrios. (2023). Llama 2: Early Adopters' Utilization of Meta's New Open-Source Pretrained Model. 10.20944/preprints202307.2142.v1.

[11] Pokhrel, Sangita & Ganesan, Swathi & Akther, Tasnim & Mapa Senavige, Lakmali Shashika Karunarathne. (2024). Building Customized Chatbots for Document Summarization and Question Answering using Large Language Models using a Framework with OpenAI, Lang chain, and Streamlit. Journal of Information Technology and Digital World. 6. 70-86. 10.36548/jitdw.2024.1.006.

[12] Gwennap, L., 2020. Groq rocks neural networks. Microprocessor Report, Tech. Rep., jan.

[13] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.T., Rocktäschel, T. and Riedel, S., 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. Advances in Neural Information Processing Systems, 33, pp.9459-9474.

[14] Zhao, Z., Monti, E., Lehmann, J. and Assem, H., 2024. Enhancing Contextual Understanding in Large Language Models through Contrastive Decoding. arXiv preprint arXiv:2405.02750.

[15] Zhang, Z., Zhao, H. and Wang, R., 2020. Machine reading comprehension: The role of contextualized language models and beyond. arXiv preprint arXiv:2005.06249.

[16] Owen, J., 2024. Optimizing AI Workflows: How'Consolidate-csv-files-from-gcs' Simplifies Data Management.

[17] Yuan, F., Yuan, S., Wu, Z. and Li, L., 2023. How Multilingual is Multilingual LLM?. arXiv preprint arXiv:2311.09071.