IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Personal Chatbot For Documents

[1]Uday Bhamre, [2]Harishankar Thakur [3]Naaz Sheikh, [4]Chandan Patil, [5] Prof. Bhaven Doshi

[1,2,3,4]UG Student, [5]Assistant Prof: Information Technology,

Trinity College of Engineering and Research, Pune,
Maharashtra, India

Abstract: This Study has been undertaken to identify issue with chatbot like ChatGPT, about Security data exposure, Sensitive data Leaks, lot of companies use chatbot, fed their data to chatbot train them and they respond to user input. In this Recent Years the Artificial Intelligence industry is evolved a lot , each industry transforming to Artificial intelligence ,in every field the Artificial intelligence is evolving Music , Image Generation, Text Generation, also the chatbot industry is most used and evolving field of the Artificial Intelligence industry , Generating Human like response to a user query is important feature of chatbot , most of companies using this chatbot to automate the Tasks and Provide Quality , Our Research mainly Focus on Tackling and increasing overall efficiency of the Chatbot and its Privacy , like ChatGPT. There are lot of Open Source Large Language Model(LLM) out there, which are Pretty Accurate and Fast and Light , with our research we are making a Personal Document Chat Bot which will allow you to Chat with PDF/TXT/CSV Document ,with 100% Privacy and Security ,no need to upload PDF over any server or online , using power of Open Source Large Language Model (LLM) we are going to achieve this Functionality, using this there will 100 percent security of you data and files will be achieved

Index Terms **– Large language model, python, Open Ai, Chatbot ,Hugging Face, Chroma Vector Store**

I. INTRODUCTION

Recently OpenAI introduced feature inside ChatGPT where user can upload Any Document like PDF , CSV ,TXT ,and ChatGPT Model will Read the data from pdf and analyse like Human and when user provide query to ChatGPT ,the model will answer to his query , any question about the PDF data ,will be answered by the Model ,this is good and Impactful feature but , w.r.t to Our Research we found there is no Data Security or Privacy to our data , this Model Learns and improves itself , so the PDF you fed the Model will use the PDF and store the data to its Knowledge Base , which not good , we need to think when we are uploading our sensitive data Documents.

Our Proposed System uses Pre-Trained Open-Source Large Language Model, to Solve this problem of Data Privacy and Allow user to Upload any Sensitive PDF or Documents and Chat With its, which will like Personal Chatbot for Documents. Our System will use Open Source Pre-Trained LLM and will fed up all the Data of Documents and the LLM will Understand the Data of Documents and Process it and user can answer.

Our first Approach is to extract data from Documents and Transform data into Such Form that Model Could Understand, LLM Model Does Not directly under text as input to it ,we are converting the extracted data ,splitting into charts and building semantics index and storing it into database(knowledge base) In this study ,we have developed a secured approach of processing Document data and processing user query , which will eliminate the need of uploading your document to external server or application and thus we will make a Personal Chatbot for Document where you can chat with Sensitive.

II. LITERATURE REVIEW

In our Research, we found there are lot of security issue with recent chatbot, "ChatGPT, developed by OpenAI in November 2022, is an AI chatbot that utilizes the Generative Pre-trained Transformer (GPT) model. OpenAI is an AI research and development company known for its innovative approaches in natural language processing. The GPT model, based on the Transformer architecture introduced by Vaswani et al. (2017), is trained on extensive datasets to generate contextually relevant and accurate responses to text-based inputs. However, as these systems become more sophisticated and widely used, concerns regarding user privacy and data protection have emerged. Large Language Models (LLMs) like ChatGPT aim to understand and generate human language, but their reliance on extensive datasets, which may contain sensitive information, raises privacy concerns. There is a risk of inadvertently capturing and exposing sensitive user data, particularly in the context of chatbots and virtual assistants where personal or confidential information is often disclosed. These concerns have been addressed in various research papers discussing the usage of LLM-based chatbots, such as those by Hariri (2023), Sebastian (2023), and Cao et al. (2023).[]
This research paper addresses the critical topic of data privacy risks associated with Large Language Models (LLMs) like ChatGPT. It acknowledges that LLMs use user data for training, which can be a threat to real privacy, especially when handling sensitive data such as financial or business documents. The paper underscores the importance of mitigating these privacy concerns through effective strategies and technologies.

The proposed techniques to ensure robust data protection in LLMs include differential privacy, federated learning, data minimization, and secure multi-party computation. Additionally, the paper explores the legal and ethical frameworks necessary for the responsible development of AI systems, considering both the potential of LLMs and the importance of user privacy. It serves as a comprehensive guide for developers, policymakers, and researchers in the field, emphasizing the need to prioritize user privacy in AI development.

The research further discusses the specific privacy and security concerns when using LLM-based Chatbots in education. It highlights the importance of robust data privacy and security policies, transparency in data collection and use, modern technologies for data protection, regular audits, and incident response plans to safeguard student data. It also emphasizes the need for educating staff and students about data privacy and security.

The paper mentions common privacy and data leakage issues with AI-based Chatbots, which can impact user trust in AI systems. These issues include unintended sharing of sensitive information, data leakage through model outputs, model extraction, and data poisoning. While the paper clarifies that models like ChatGPT don't have access to personal data, it warns about potential risks if communication channels are not secure.

In addition to these considerations, the paper highlights research discussing the use of pre-trained foundation models (PFMs) like GPTs in edge intelligence for AI services in the Metaverse. It proposes a framework for efficient resource management and introduces the "Age of Context" metric to balance latency, energy consumption, and accuracy in mobile AI services within the Metaverse.

Future research scopes include investigating the cybersecurity, data privacy, and ethical implications of increased AI adoption in the Metaverse, as well as exploring the broader impact of AI technologies in this emerging virtual environment. These areas of study are crucial as the Metaverse continues to evolve and expand.
i] Unintended Sharing of Sensitive Information: This occurs when a user unknowingly shares personal or sensitive data with the AI system (Sweeney, L., 2002). For instance, a user might share their credit card information, believing that the AI is secure. While non-PII AI models like ChatGPT do not have the ability to recall or store this information, given it stores non-PII information temporarily to improve performance for 30 days, the data could potentially be intercepted during transmission if the communication channel is not secure.

ii) Data Leakage through Model Outputs: Even though LLM models like ChatGPT do not know specifics about the data they were trained on, they can sometimes generate outputs that seem to refer to specific data or reveal sensitive information. However, these outputs are generated based on patterns learned during training and do not reflect access to any specific data sources or confidential databases. The AI could

"hallucinate" specific, sensitive-looking details in responses. The model is not leaking real-world sensitive data that it learned during training—it's making things up based on the patterns it learned

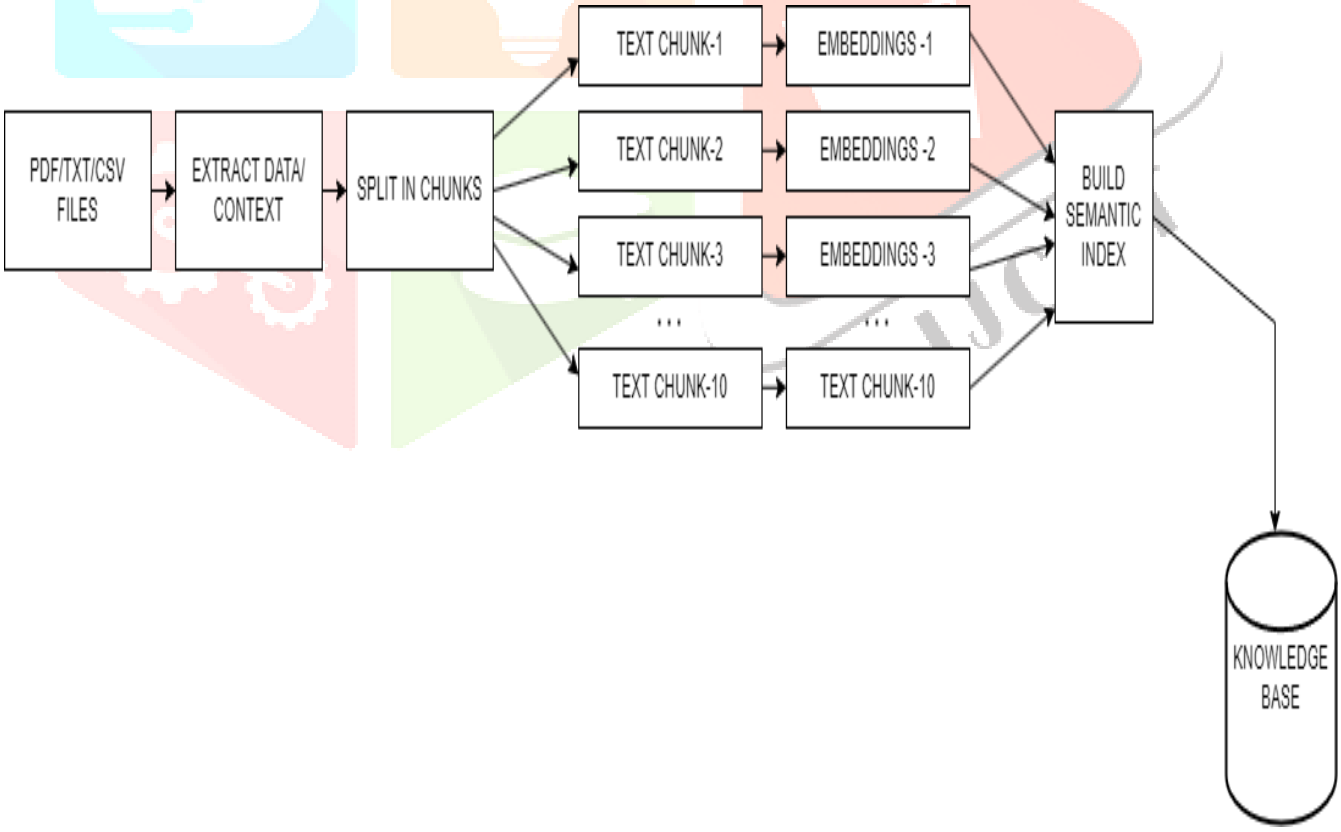
iii) Model Extraction: This involves an attacker using the outputs of a machine learning model to create a copy of that model without access to the original training data. If successful, the attacker could use the extracted model for malicious purposes, potentially undermining the security and integrity of the original system

iv) Data Poisoning: This is a type of attack where the attacker introduces harmful data into the model's training data with the aim of influencing its future predictions or behaviour. It's a significant threat for systems that continually learn from their interactions with users

This research discusses the use of pre-trained foundation models (PFMs) like generative pre-trained transformers (GPTs) in edge intelligence to provide AI services.

III. PROPOSED SYSTEM ARCHITECTURE

Previous System was taking Document as input and uploading to server and returning response Our Proposed System uses Pre-Trained Open-Source Large Language Model, to Solve this problem of Data Privacy and Allow user to Upload any Sensitive PDF or Documents and Chat With its, which will like Personal Chatbot for Documents. Our System will use Open Source Pre Trained LLM and will fed up all the Data of Documents and the LLM will Understand the Data of Documents and Process it and user can answer.



Extracting Data from Documents and Storing it in Knowledge base

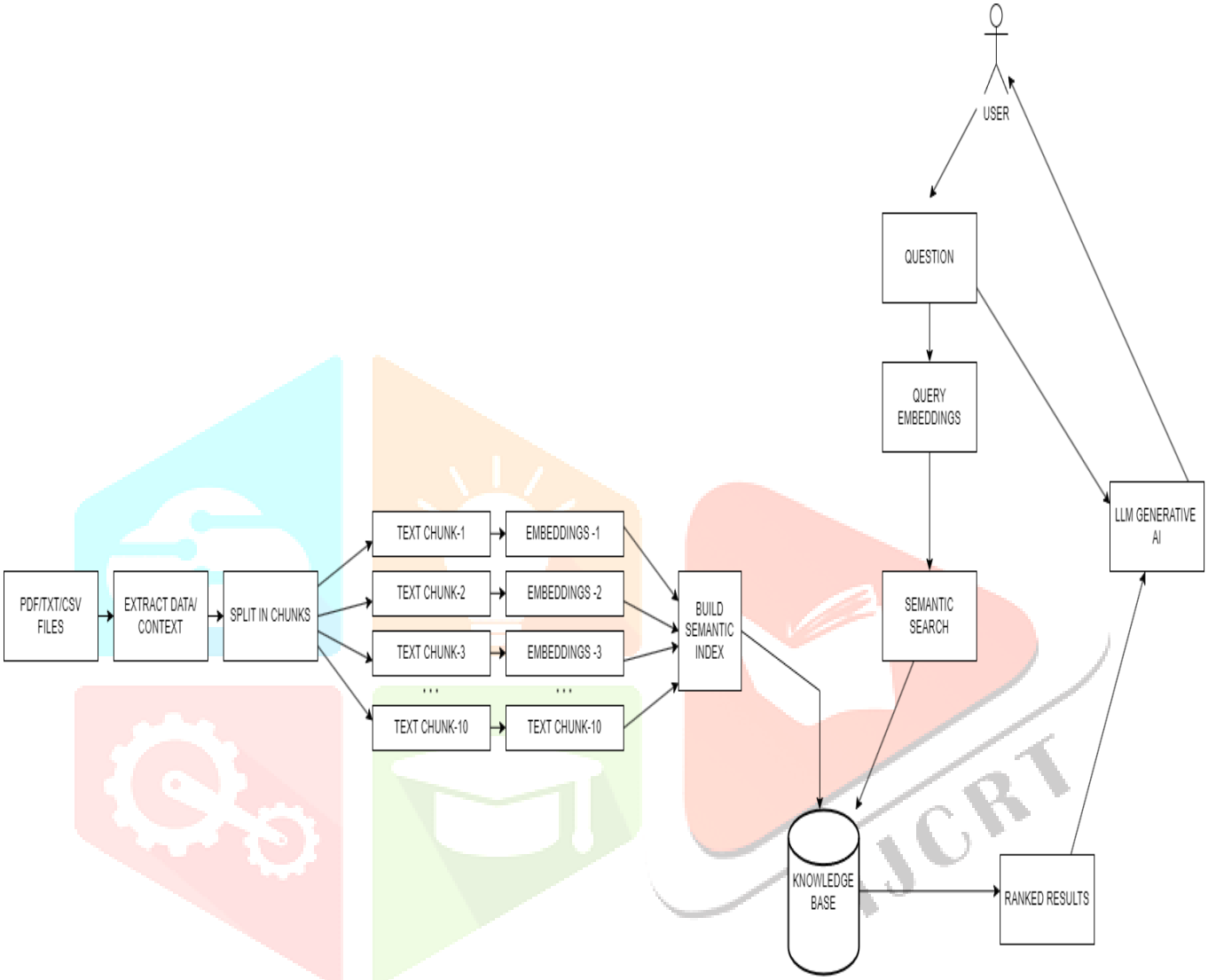


Figure No.1.0 System Architecture

IV. CONCLUSION

The Previous System of Chat with Document was not Secured and leading to sensitive data exposure and leaks, person was not able to upload or chat with personal document like Balance Sheet, Personal Documents etc, which was major flaw, the peoples with this document was not able to take advantages of features, they all need to do manual analysing od Documents

Our Proposed system tackles this problem all allows user to upload and chat with documents with full Privacy and Secured manner, no data leaves the Device of user, this system uses Open source LLM for Generating Human like Responses to User Document , in First Step we will extract and Store user Data from Documents and later the LLM Model will have Access to the Data and the LLM Model will be Pre trained and the Model will analyse the user query and return most Ranked response to user , main advantage of our system is it will be completely offline will be in users device

V. REFERENCES

[1] Privacy and Data Protection in ChatGPT and Other AI Chatbots: Strategies for Securing User Information Glorin Sebastian, Georgia Institute of Technology, USA* https://orcid.org/0000-0003-2543-9127

[2] ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope Partha Pratim Ray Sikkim doi.org/10.1016/j.iotcps.2023.04.0

[3] A Comprehensive Study of ChatGPT: Advancements, Limitations, and Ethical Considerations in Natural Language Processing and Cybersecurity Moatsum Alawida 1, *, Sami Mejri 2, Abid Mehmood 1, Belkacem Chikhaoui 3, * and Oludare Isaac Abiodun 4 doi.org/10.3390/info14080462

[4] ChatGPT for good? On opportunities and challenges of large language models for education Enkelejda Kasneci a,* , Kathrin Sessler a , Stefan Küchemann b , Maria Bannert a , Daryna Dementieva a , Frank Fischer b , Urs Gasser a , Georg Groh a , Stephan Günnemann a , Eyke Hüllermeier b , Stephan Krusche a , Gitta Kutyniok b , Tilman Michaeli a , Claudia Nerdel a , Jürgen Pfeffer a , Oleksandra Poquet a , Michael Sailer b , Albrecht Schmidt b , Tina Seidel a , Matthias Stadler b , Jochen Weller b , Jochen Kuhn b , Gjergji Kasneci c

[5] IndiaAiyappa, R. (2023). Can we trust the evaluation on ChatGPT? arXiv preprint arXiv:2303.12767 Akhawe, D., Amann, B., Vallentin, M., & Sommer, R. (2013, November). Here's my cert, so trust me, maybe? understanding TLS errors on the web. ACM.

[6] Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: Implications in scientific writing. Cureus, 15, 2. doi:10.7759/cureus.35179 PMID:36811129

[7] Balebako, R., Marsh, A., Lin, J., Hong, J., & Cranor, L. (2014). The Privacy and Security Behaviors of Smartphone App Developers. USEC. doi:10.14722/usec.2014.23006

[8] Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. Pattern Recognition, 84, 317–331. doi: 10.1016/j.patcog.2018.07.023

[9] Cao, Y. (2023). A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt. arXiv preprint arXiv:2303.04226