**DEVASHISH PANIGRAHI**

**Abstract-**The research objective is to explore the integration of AI and Blockchain technologies for enhanced data management and secure, decentralized storage. This includes reshaping file analysis within a secure environment. The implementation involves an AI-powered chat module using the advanced GPT-4 model to facilitate efficient file analysis and interaction. Additionally, the introduction of the innovative Chroma Vector Database aims to optimize data retrieval and analysis within Blockchain storage. The theoretical exploration envisions a future with secure, efficient, and intelligent file analysis, addressing challenges in complex files understanding. The potential applications span across fields such as research, cyber forensics, and various industries, aiming to provide solutions for complex files understanding challenges. (To MAKE IT MORE ATTRACTIVE)

**Keywords -** GPT-4**,** LangChain Layer**,** Chroma Vector Database**,** Blockchain**,** AES**.**

**1. INTRODUCTION**

In the ever-evolving landscape of technology, the confluence of Artificial Intelligence (AI) and Blockchain stands at the forefront of innovation, offering unprecedented possibilities for data management, security, and user interaction. As the digital era continues to unfold, the significance of effective data governance cannot be overstated, with organizations and individuals relying on robust mechanisms for analysis and safeguarding sensitive information [1-2]. Blockchain technology, renowned for its decentralized and immutable ledger, emerges as a beacon in the realm of data storage and security. Its transformative capabilities have positioned it as a stalwart solution, ensuring transparency, integrity, and resilience in the face of evolving cyber threats. Simultaneously, the field of AI has undergone remarkable advancements, particularly in natural language processing, paving the way for novel applications that promise to reshape the landscape of human-computer interaction [3-8]. This research paper embarks on a pioneering exploration at the intersection of these influential domains, delving into the fusion of AI and Blockchain to redefine fundamental aspects of file analysis and user interaction paradigms. The focal point of this investigation lies in the implementation of an AI-enabled chat module within a Blockchain storage environment, leveraging the cutting-edge capabilities of the GPT-4 language model. The primary goal is twofold: to optimize file analysis processes and to create a secure, decentralized interface that empowers users to interact intuitively with their data. An innovative facet of this research involves the incorporation of the Chroma Vector Database, presenting a novel approach to data representation and retrieval. This addition aims to enhance file analysis efficiency and expedite data retrieval within the Blockchain, effectively addressing existing limitations in traditional data storage and retrieval methodologies [10-14]. Through a comprehensive exploration of theoretical foundations, technological integration intricacies, and potential applications, the research sheds light on the transformative potential of AI-enabled file analysis within a Blockchain storage ecosystem. Expanding its scope further, the paper integrates AI-enabled audio processing within the proposed chat module, thereby broadening the spectrum of intelligent interaction in the Blockchain storage environment. Capitalizing on GPT-4's advanced natural language understanding, the research endeavors to interpret spoken commands seamlessly. Moreover, the Chroma Vector Database is extended to support the storage and retrieval of audio data, presenting a holistic approach that envisions a unified platform where users seamlessly engage with both textual and auditory data in a secure and decentralized manner. This holistic approach envisions a unified platform where users seamlessly engage with both textual and auditory data in a secure and decentralized manner. The research meticulously examines the challenges and opportunities arising from the fusion of AI, Blockchain, and audio processing, thereby envisaging a future where AI-driven chat modules comprehend both textual and voice interactions [16-18]. This foresight enriches the versatility of data interaction within the decentralized ecosystem, setting the stage for a transformative paradigm in the evolving digital landscape.

**2. OBJECTIVES**

We aim to revolutionize file analysis across domains by creating an AI-driven system that enhances efficiency and accuracy, particularly in legal file verification and crime detection. This innovation extends to integrating secure Blockchain technology for decentralized storage, ensuring data integrity and privacy exemplified in the recruiting industry for streamlined resume analysis. Augmenting human-computer interaction, a user-friendly chat module powered by the advanced GPT-4 language model is introduced, applicable for research file comprehension and educational efficiency. The system also prioritizes efficient data retrieval through the cutting-edge Chroma Vector Database, showcasing its effectiveness in managing and accessing data in complex business files for companies. With a versatile design, this solution transcends industries, finding applications in healthcare, finance, and cybersecurity, facilitating data-driven decision-making, efficient financial analysis, and superior customer support.

**3.CONTRIBUTION**

In the dynamic tech landscape, the synergy of AI and Blockchain redefines data management. This research focuses on implementing a GPT-4-powered AI chat module in Blockchain storage, optimizing file analysis and ensuring secure, decentralized user interaction, with integration of the Chroma Vector Database to enhance data retrieval. Envisioning a future where AI comprehends both textual and voice interactions, this study pioneers a transformative paradigm in the decentralized ecosystem. Section I introduces the research's core focus on AI and Blockchain integration, emphasizing its potential for revolutionizing data management. Section II succinctly reviews the synergy of Blockchain and AI, highlighting their impact on secure data storage and human-computer interaction. Section IV serves as a comprehensive exploration and synthesis of existing knowledge and research pertaining to the subject matter at hand. Our innovative file analysis model comprises five key layers. Section V details the components of our file analysis model, emphasizing their roles in optimizing data management within Blockchain. In section VI, we present the proposed architecture for the AI-enabled interaction system aimed at facilitating file analysis within a Blockchain storage environment. Additionally, we delve into the results and comparative analysis of the LLM model and NLP model. Section VII contains conclusion and future scope.

**4. SYSTEMATIC LITERATURE REVIEW**

In the landscape of document understanding and natural language processing, several pioneering studies have emerged, each contributing significantly to their respective domains. Wang et al. [1] introduced the Visually Rich Document Understanding (VRDU) benchmark, emphasizing the extraction of structured data from complex business documents. This benchmark addresses challenges such as hierarchical entities and diverse templates, setting a standard for future research. Concurrently, Yang et al. [2] presented CASIA-onDo, an online database for handwritten texts, facilitating layout analysis tasks. Chew et al. [3] streamlined deductive coding in qualitative research with LLM-assisted content analysis (LACA), demonstrating GPT-3.5's competence in matching human coder agreement levels. Expanding on language models, Karpinska et al. [4] explored paragraph-level translation with GPT-3.5, showcasing higher-quality translations across multiple language pairs. Cui et al. [5] delved into the emerging field of "Document AI," utilizing deep learning for automatic understanding of corporate documents. They highlighted challenges such as OCR integration and data quality variations. Wu et al. [6] examined ChatGPT's language-centric limitations, leading to the development of Visual ChatGPT, incorporating visual foundation models for a richer interactive experience. Addressing the integration of human ingenuity and machine memory in biomedical science, Zhang et al. [7] introduced GeoGPT, leveraging LLMs for geospatial tasks, showcasing its potential in diverse GIS applications.

Jeong et al. [8] introduced generative AI services using LLMs, focusing on information scarcity challenges and proposing the Retrieval-Augmented Generation (RAG) model. Arawjo et al [9] introduced ChainForge, a visual toolkit for evaluating LLMs, providing a user-friendly interface for prompt engineering and hypothesis testing. Zhang et al [10] addressed limitations in log-based anomaly detection methods, proposing the use of ChatGPT for more efficient and interpretable anomaly detection. Harper et al. [11] explored the transformative potential of LLMs in legal and regulatory compliance, integrating LangChain, GPT-4, and text embeddings for reliable domain-specific outputs. Bhavan Jasani [12] introduced DocFormer, a multi-modal transformer architecture for Visual Document Understanding (VDU). Kaiming et al. [13] introduced a residual learning framework, achieving deeper neural network architectures with improved optimization and accuracy. Jianhua et al. [14] detailed the challenges and contributions of the SROIE competition, emphasizing its significance in document analysis systems. Mathew et al [15] introduced DocVQA, a dataset for Visual Question Answering, highlighting the performance gap between existing models and human accuracy. Powalski et al. [16] tackled the Natural Language Comprehension challenge using the TILT neural network architecture, achieving state-of-the-art results. Lev Ratinov and Dan Roth[17] delved into key design challenges in developing a robust Named Entity Recognition system, culminating in an NER system boasting a remarkable 90.8 F1 score. Graliński et al [18] addressed the scarcity of benchmarks for Key Information Extraction (KIE) tasks, presenting two novel datasets and benchmarking against various models. Lianwen et al. [19] proposed LiLT, a language-independent layout transformer demonstrating competitive performance across eight languages. Fei Wu et al [20] addressed the interdependence of text reading and information extraction, proposing a unified end-to-end network for improved efficiency and accuracy.

Florencio et al [21] introduced LayoutLMv2, incorporating novel pre-training tasks and achieving significant performance improvements across visually-rich document understanding tasks. In summary, these studies collectively contribute to advancing document understanding, natural language processing, and the utilization of large language models, showcasing innovative approaches, benchmark datasets, and improved system performances. (PLEASE CHECK REFERNCES)

**5. PROPOSED MODEL**

Our proposed model represents a cutting-edge approach to file analysis, leveraging AI-enabled interaction across five distinct layers for unparalleled functionality and security. Embarking on a journey towards a seamlessly integrated data management solution, our proposed model encompasses various intricate layers to ensure both security and efficiency. To initiate this transformative process, users will experience a user-friendly interface that serves as a gateway to a world of enhanced file management. Below is the process of the working on our model.

The user initiates a file upload through the UI layer, which seamlessly forwards the files to the Encryption Layer. Subsequently, the Encryption Layer encrypts the files and transmits the data to the Blockchain Storage Component. This component diligently records the transaction on the Blockchain and securely stores the encrypted files. In a parallel user interaction, the individual enters a natural language query via the UI layer, which promptly dispatches the query to the Chat Module. Within the Chat Module, the Query Vectorization process transforms the user's input into a vector, initiating the Vector-Based Comparison in the Chroma Vector Database to identify relevant files. Upon identification, the Natural Language Generation kicks in, crafting a response based on the relevant files. Finally, the Chat Module efficiently sends the generated response back to the UI layer, completing the seamless interaction loop. Fig. 1 shows the overall working of the model.

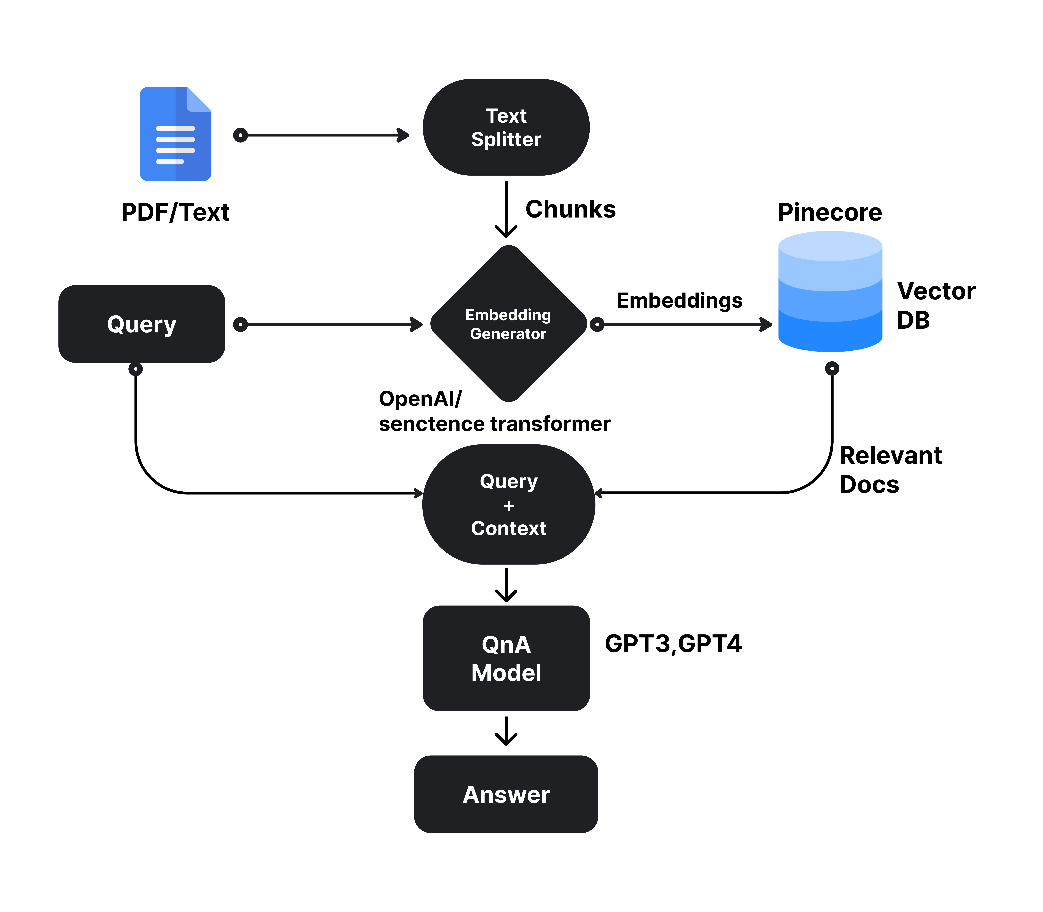


Fig. 1. Flow chart of the proposed Model

At the forefront is the UI layer, designed to facilitate intuitive and user-friendly interaction. This layer serves as the gateway for users to input queries, access analysis results, and navigate through the application seamlessly. Ensuring the utmost security of user data, the AES blockchain layer comes into play, employing advanced encryption techniques to safeguard the files provided by users. By integrating blockchain technology, this layer not only enhances data security but also ensures transparency and integrity throughout the file analysis process. In parallel, the vectordatabase layer stands as the repository for storing files in chroma vectors. This innovative approach to storage enables efficient indexing and retrieval of files, optimizing the overall performance of the system. The LangChain layer serves as a vital link between user queries and the documents stored in the database. Leveraging sophisticated language processing techniques, this layer interprets user queries, retrieves relevant documents, and presents the analysis results coherently. At the heart of the model lies the GPT layer, which acts as the brain of the entire application. Harnessing the power of GPT-4 and advanced algorithms, this layer maps the vectors extracted from files, performs insightful analysis, and generates actionable insights for users. Together, these five layers synergize to form a robust, efficient, and secure file analysis solution, poised to revolutionize the way users interact with and analyze their data. Fig.2 shows the 5 layers of our proposed model.

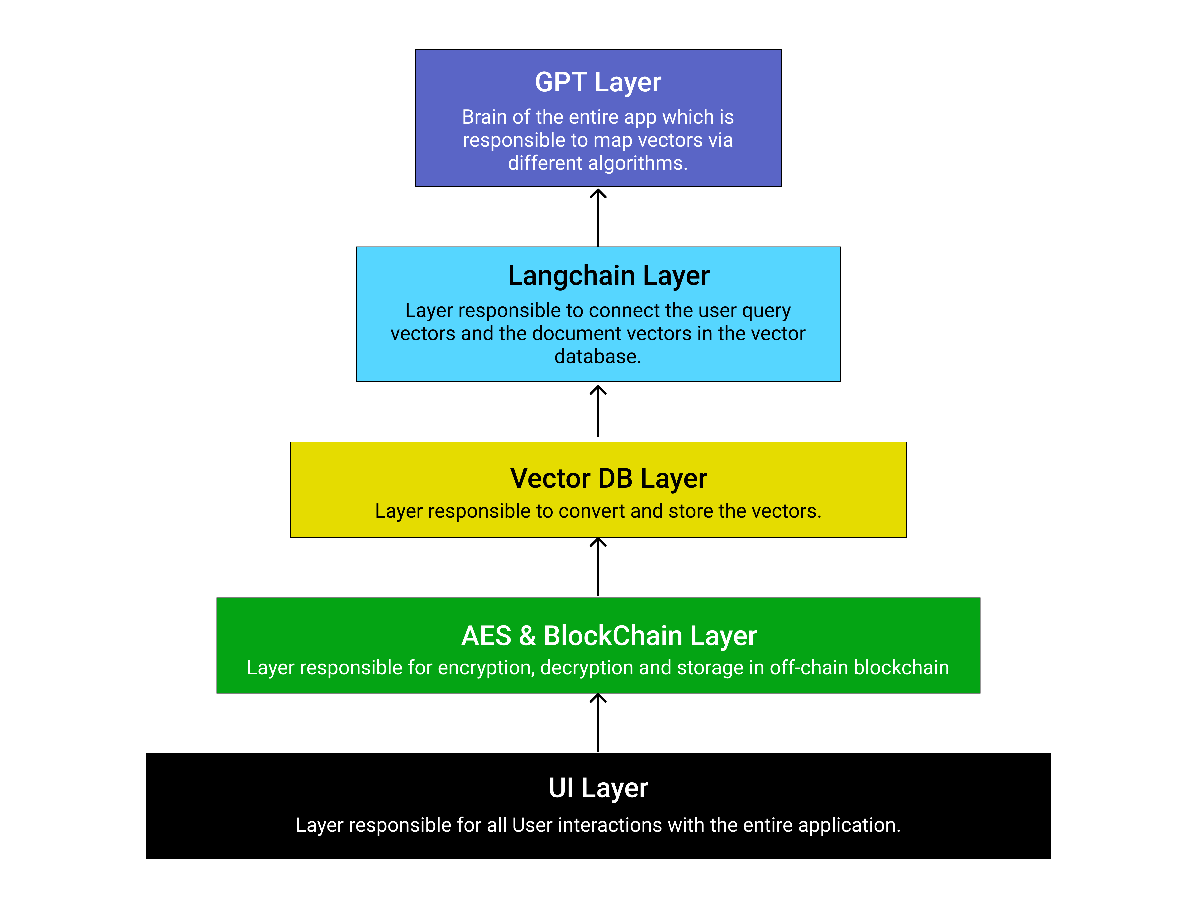
****

Fig.2. Five layers of Proposed model

**4.1 Five Layers of Proposed Model:**

In the ever-evolving landscape of artificial intelligence, the proposed model stands as a testament to innovation and efficiency. Comprising five meticulously designed layers, this model embodies a holistic approach to solving complex problems and pushing the boundaries of AI capabilities. Each layer plays a crucial role in the overall architecture, contributing to the model's robustness and adaptability

**4.1.1 UI Layer: User Interaction and Document Upload**

The first layer of the architecture is the User Interface (UI) layer as shown in Fig.3, where users interact with the system. This layer serves as the entry point for users to upload documents. Key functionalities and components within this layer include:

User Interaction: Users interact with the system through a user-friendly interface, allowing them to initiate various actions such as document upload, queries, and preferences. Document Upload: Users can upload one or more documents through the UI. These documents are subsequently processed, encrypted, and prepared for storage in the Blockchain layer.

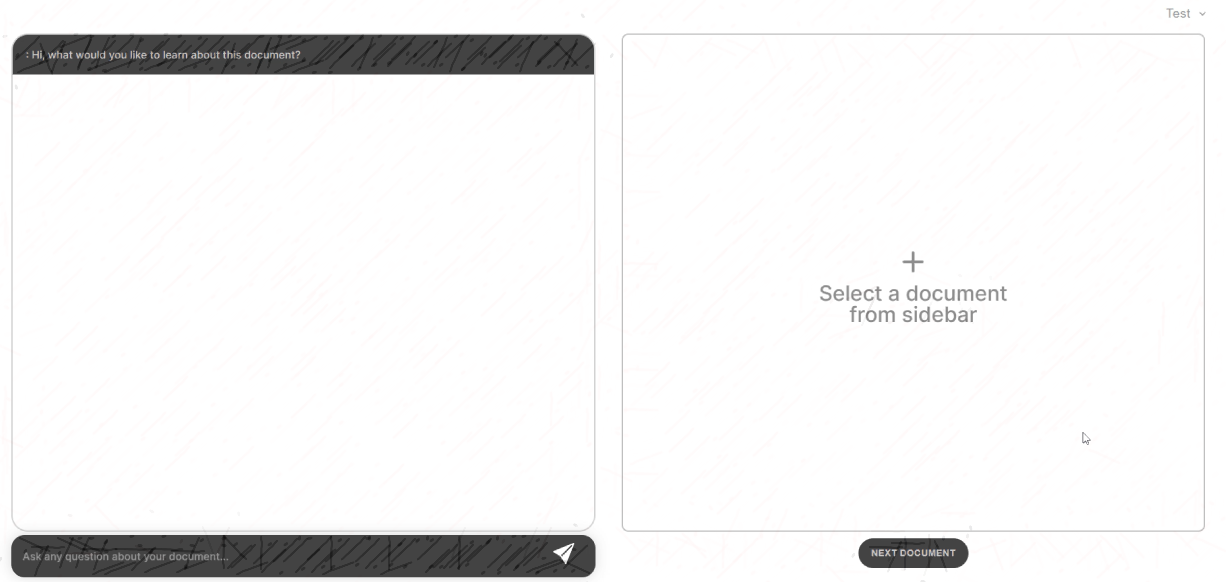


Fig.3. Graphical User Interface (GUI) of Our Proposed Model

**4.1.2 Encryption Layer: Secure Document Storage in Blockchain:**

After document upload, the next layer focuses on ensuring the security and immutability of documents through encryption and storage in the Blockchain. Key functionalities and components in this layer include:File Encryption: Uploaded documents are encrypted using robust cryptographic algorithms to protect data integrity and privacy. Blockchain Storage: Encrypted documents are stored on the Blockchain. The Blockchain layer records the transaction details and ensures decentralized and tamper-proof storage. Smart Contracts: Smart contracts may be employed to manage access control and permissions for uploaded documents, enhancing security further. Fig.4 shows how the uploaded document is stored in the database after securely encrypted using Blockchain.

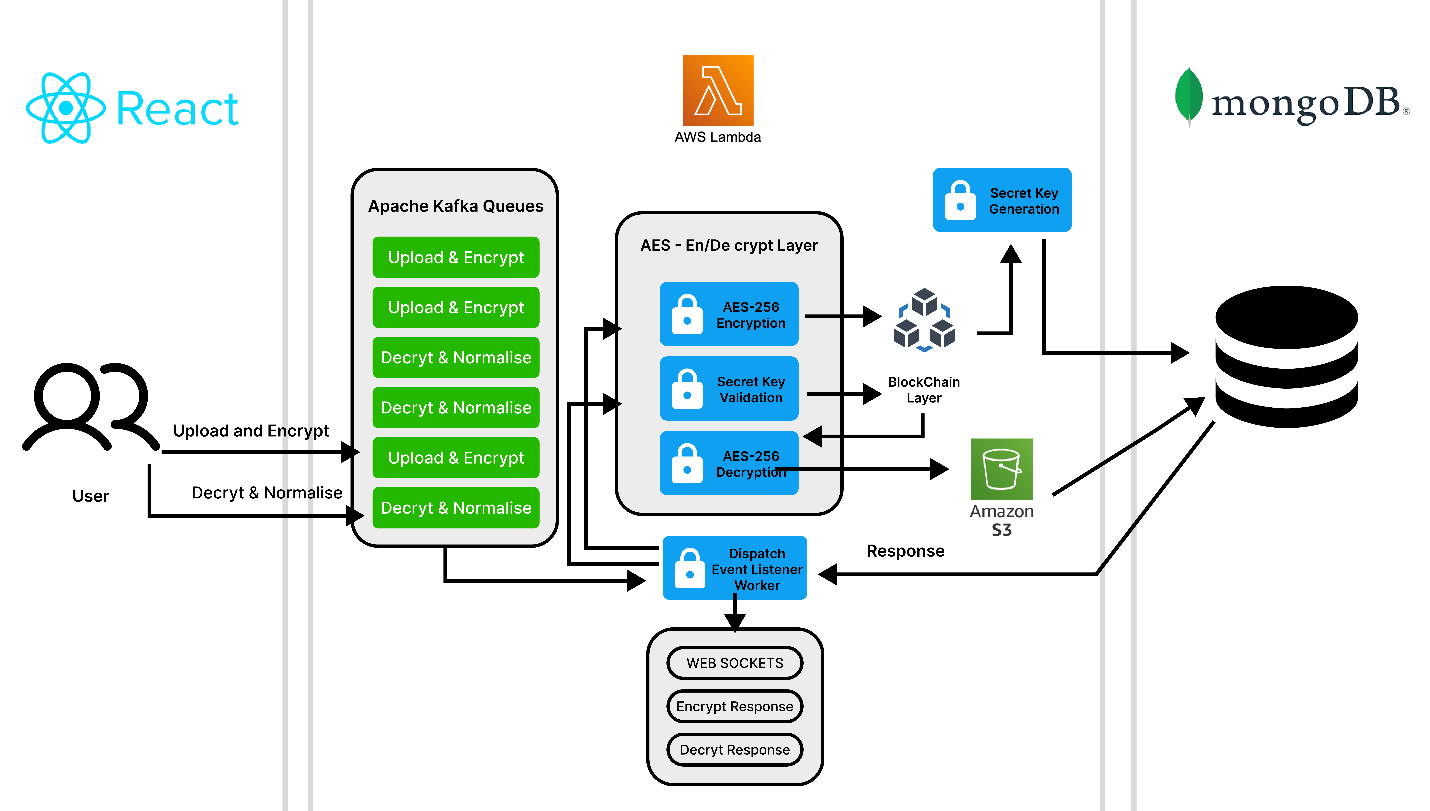


Fig.4. Secure Document Storage in Blockchain

**4.1.3 Store and Conversion to Vectors: Utilizing Chroma Vector Database:**

The third layer focuses on converting document content into vectors and efficiently storing them in the Chroma Vector Database for subsequent analysis. Key functionalities and components in this layer include: Content Vectorization: Upon user request or system triggers, the content of uploaded documents is converted into vectors using the Chroma Vectorization technique. Chroma Vector Database: Vectors are stored in the Chroma Vector Database, which is designed for efficient storage, retrieval, and similarity-based searching. Fig.5 shows the chroma vector database containing chunks of uploaded file to analyze efficiently and interact with queries.

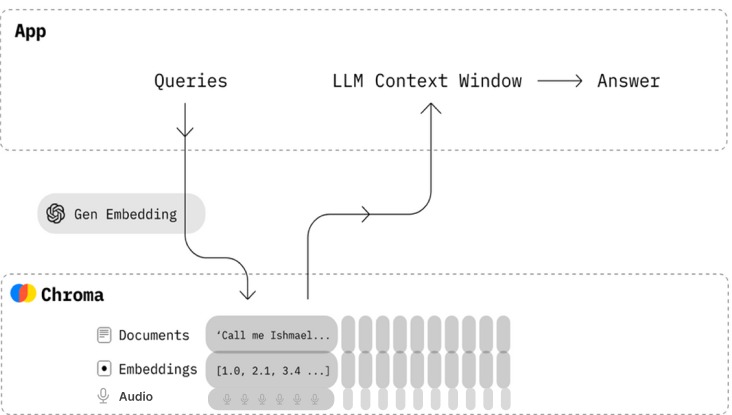
Metadata Management: Alongside vectors, metadata related to documents (e.g., timestamps, user information) is stored to facilitate document tracking and management.

Fig. 5. **Efficient Data Repository Layer for storing vectors**

**4.1.4 Chat Module: User Interaction and Question-Answering**

The fourth layer introduces the chat module, enabling users to interact with the system through natural language queries and receive intelligent responses based on the analyzed documents as shown in the Fig. 6. Key functionalities and components in this layer include: User Queries: Users can ask questions or request information through natural language queries within the chat interface. Vectorization of Queries: User queries are converted into vectors to facilitate efficient comparison with document vectors. Vector-Based Comparison: The system compares user query vectors with document vectors stored in the Chroma Vector Database to retrieve relevant documents and responses Natural Language Generation: Responses are generated using GPT technology, which leverages the underlying GPT layer to ensure coherent and context-aware answers.

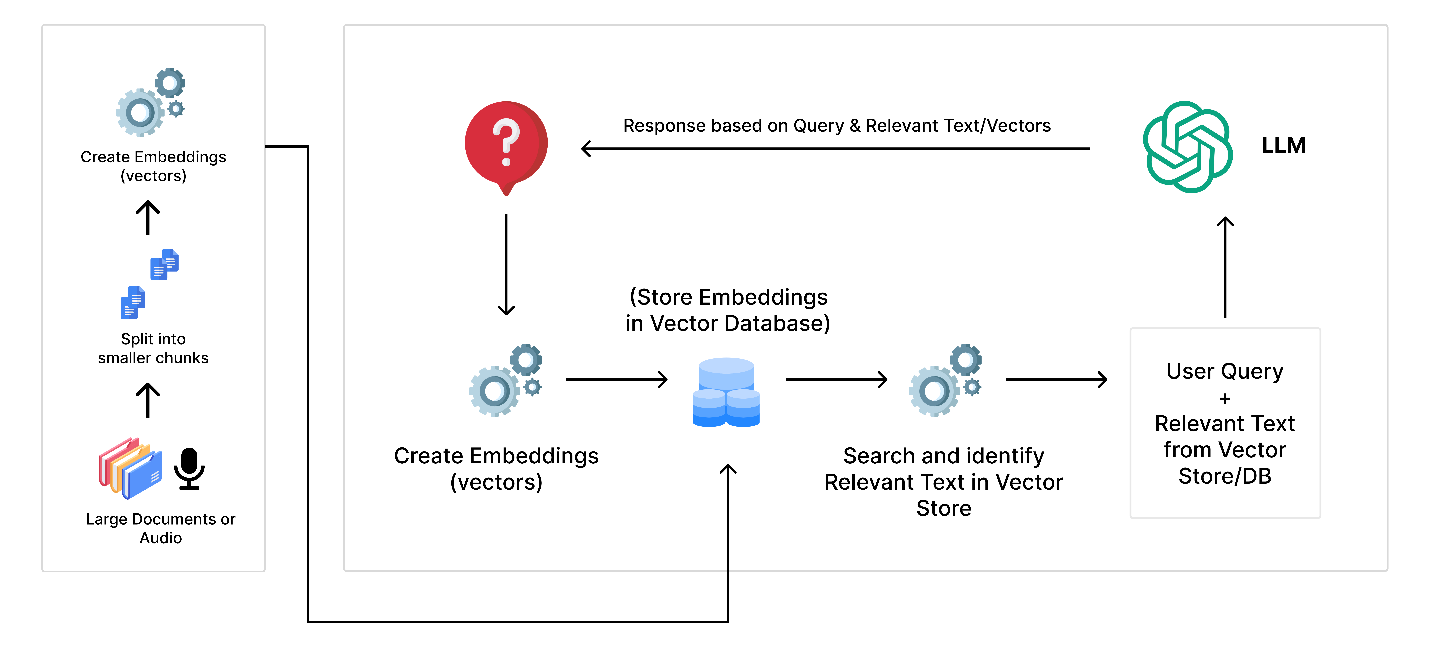


Fig. 6. Conversion of query vector and document vector in vector database

**4.1.5 Underlying GPT Layer: Natural Language Understanding and Generation:**

The fifth and underlying layer is powered by GPT technology, responsible for understanding user queries, interpreting document vectors, and generating meaningful responses. Key functionalities and components in this layer include: Natural Language Understanding: GPT models process user queries, interpreting intent, context, and semantics. Vector-Text Mapping: GPT models understand and map vectors to meaningful text, aiding in contextual analysis. Response Generation: GPT models generate responses to user queries, ensuring coherent, context-aware, and informative answers. Feedback Loop: The GPT layer may incorporate feedback mechanisms to improve response quality and system performance over time. This proposed architecture seamlessly integrates user interaction, document security, vector-based analysis, natural language understanding, and response generation to create an AI-enabled interaction system capable of intelligent file analysis within a Blockchain storage environment. Each layer plays a crucial role in ensuring the system's functionality, security, and user-friendliness. The Fig. 7 gives a pictorial representation of how the underlying GPT layer works.

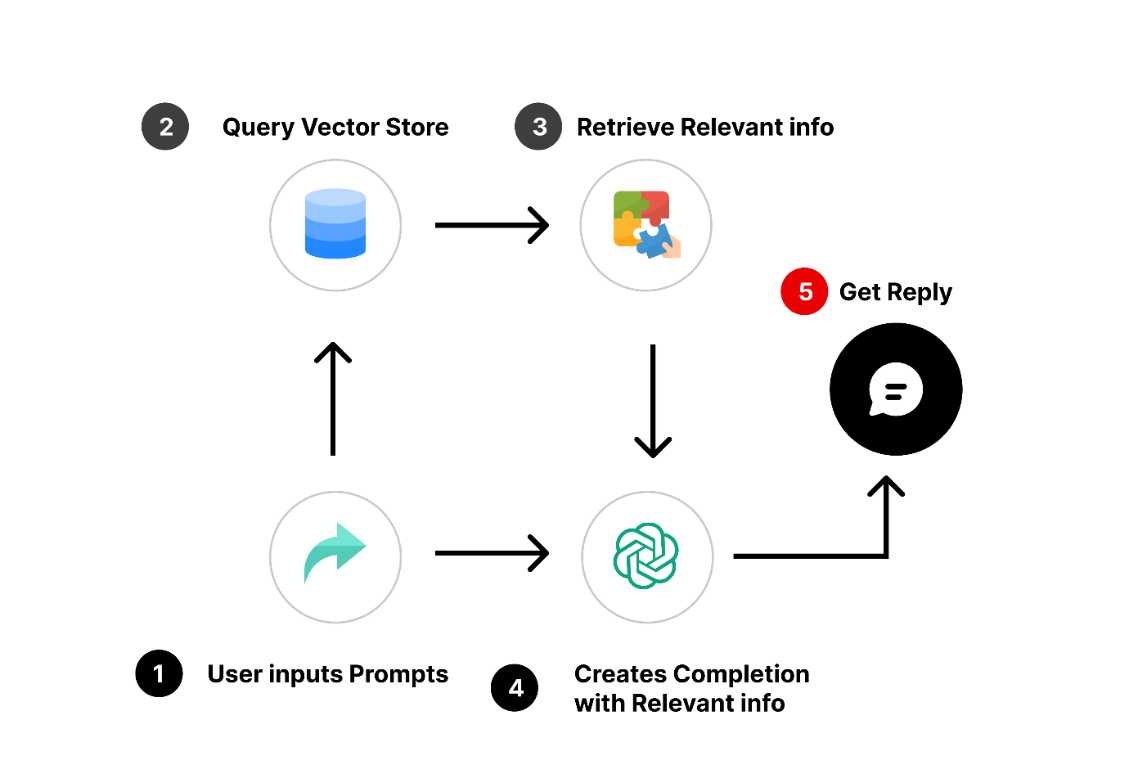


Fig. 7. Mapping vectors according to the query

Our model features a layered architecture for efficient data management. Users interact through intuitive interfaces like the User Interface and Document Upload Interface. Documents are encrypted and stored on a Blockchain network for security. They're then converted into vectors for efficient storage and representation. The Chat Module allows natural language interaction, processing user queries and generating responses. The underlying GPT Layer interprets queries, generates responses, and adapts through user feedback. This architecture ensures a seamless user experience, blending security, efficiency, and natural language interaction effectively. Fig. 8 shows the different components of the proposed model which has been integrated to work efficiently.

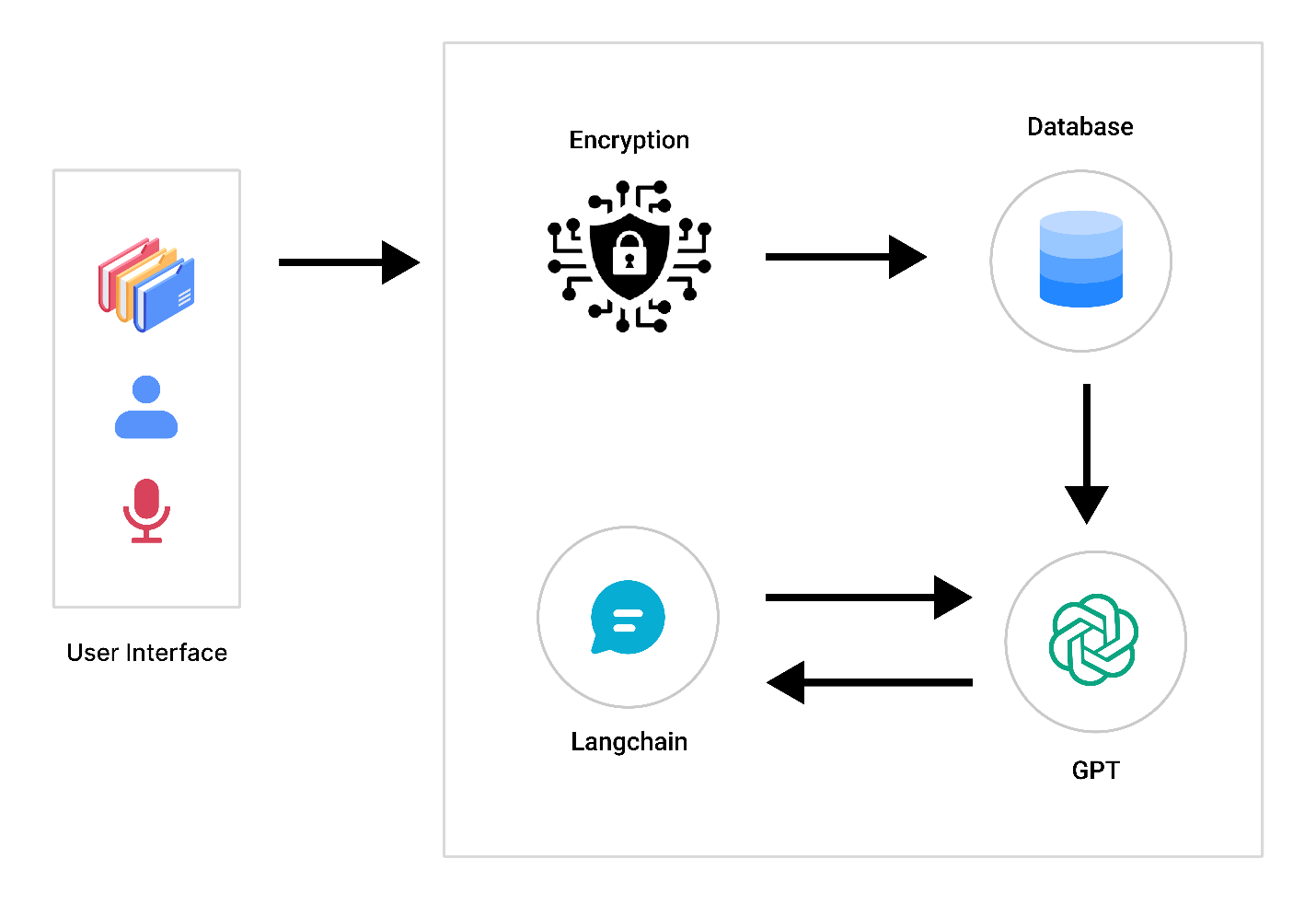
****

Fig. 8. Component Diagram of our proposed model

**6. RESULT AND DISCUSSION**

Our paper reveals that LLMs demonstrate superior performance in context understanding and task execution compared to traditional NLP models. Our analysis contrasts traditional NLP models against LLMs like GPT-3, examining tasks, training data, architectural frameworks, depth of context understanding. Traditional models are task-specific and limited context comprehension, while LLMs leverage vast pre-trained data for richer contextual insights.Fig.9 is the Graphical User Interface of our Model, where the user is going to upload the document or audio by clicking on the ‘Plus’ icon.

A screenshot of a computer

Description automatically generated

Fig. 9. UI Interface of our proposed model

The depicted Fig.10 illustrates the pivotal step in the process wherein users select the document or audio file intended for subsequent analysis.

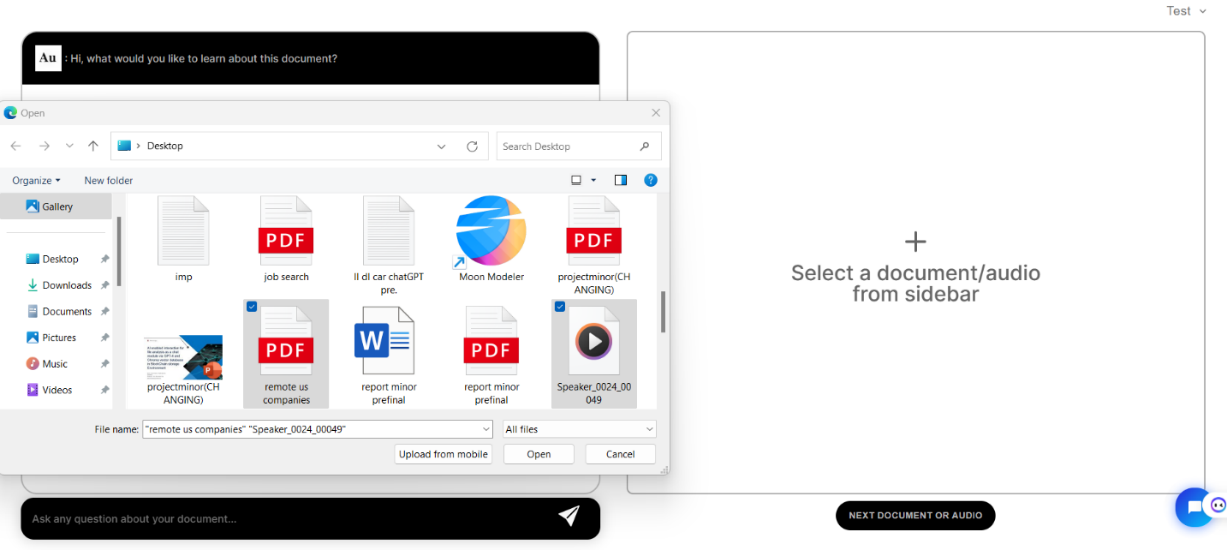


Fig. 10. Uploading a document/audio

A screenshot of a computer

Description automatically generatedFig.11 illustrates the selected document positioned on the right side of the interface, showcasing the contents of the document. At this juncture, the user is poised to engage the model with queries and inquiries pertinent to the subject matter. This professional setup ensures an organized and efficient interaction, facilitating seamless communication and information retrieval.

Fig. 11. Asking a query

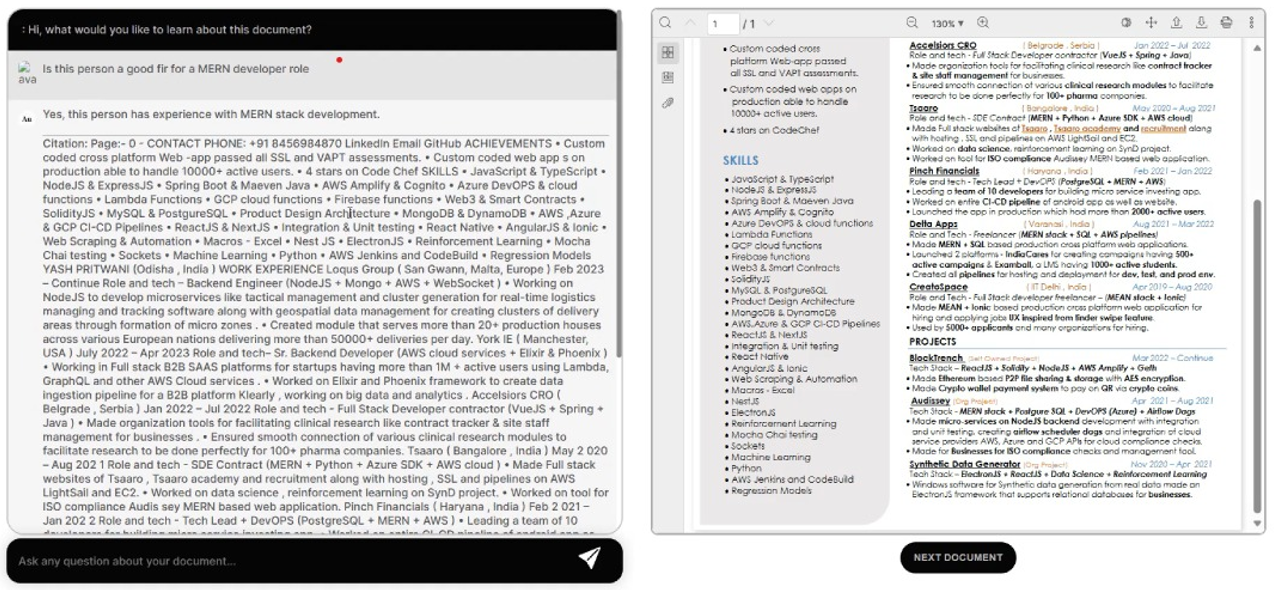
Fig. 12 depicts the response elicited from the LLM agent in response to the posed inquiry

Fig. 12. Answer upon the question asked to LLM agent

Write 4-5 lines summarize

**6.1. COMPARATIVE ANALYSIS**

1. Architecture  
   NLP Models has Enhanced Linguistic Understanding i.e. NLP models are designed to understand and process human language, enabling computers to interpret, analyse, and generate text in a way that is meaningful to users. These models are applied to various task-specific applications, such as sentiment analysis, named entity recognition, machine translation, and text summarization, among others. Moreover, NLP models employ vectorization techniques to represent textual data numerically. Common methods include one-hot encoding, word embeddings and more recently, contextual embeddings from models like BERT. Traditional NLP models often use architectures such as Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs), Convolutional Neural Networks (CNNs), and other structures tailored to specific tasks. Compared to Large Language Models (LLMs) like GPT-3, traditional NLP models may have limitations in capturing extensive contextual dependencies, as they are typically trained on task-specific datasets and lack exposure to diverse and large-scale language.Fig.14. Illustrates the working of NLP Model.

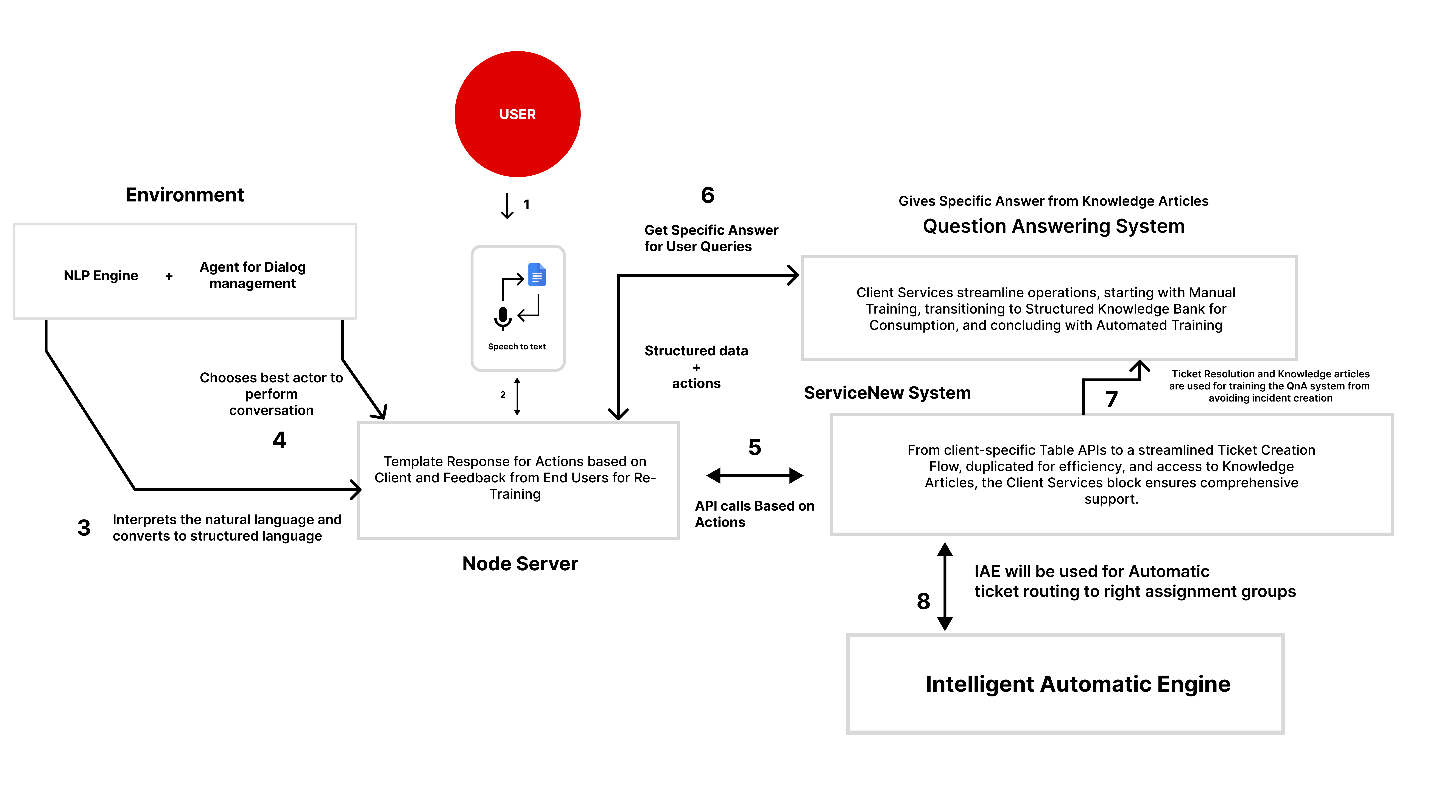


Fig. 14. Working of NLP model

LLMs, such as GPT-3, typically use the transformer architecture and are pre-trained on vast and diverse datasets containing a wide range of text from the internet. LLMs generate contextualized embeddings for each token in a sequence and excel at transfer learning, where knowledge gained during pre-training on a large corpus can be fine-tuned for specific tasks with smaller datasets. Moreover LLMs are capable of natural language generation, allowing them to produce coherent and contextually relevant text. Fig.15 illustrates the working of LLM model.

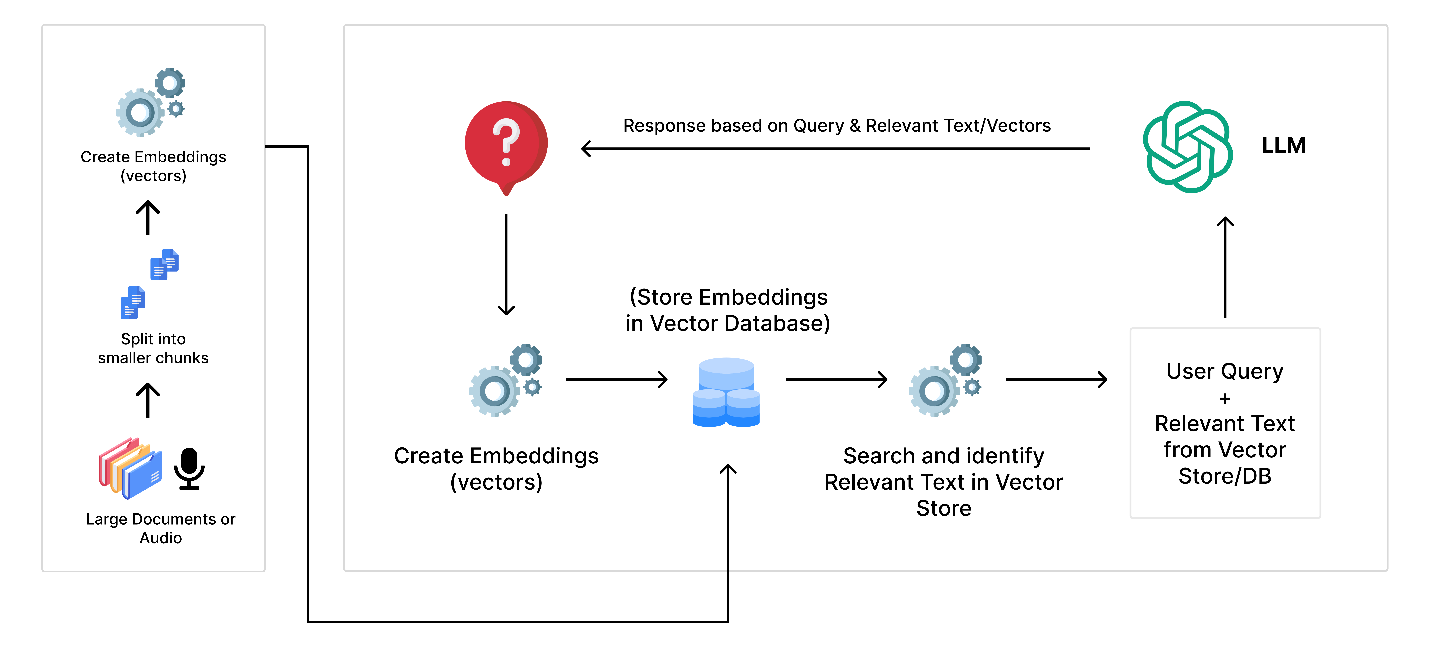


Fig. 15. Working of LLM model

Table 2. Comparison of NLP and LLM model

|  |  |  |
| --- | --- | --- |
| **Aspect** | **NLP Model** | **LLM(eg GPT-3)** |
| Task | Performs specific NLP tasks like sentiment analysis, named entity recognition, etc. | General-purpose language understanding and generation. |
| Training Data | Typically trained on task-specific datasets. | Pre-trained on vast amounts of diverse text data. |
| Architecture | Various architectures (e.g., RNNs, LSTMs, GRUs, CNNs) depending on the task. | Transformer architecture is commonly used. |
| Context Understanding | Limited context understanding based on the task. | Captures rich context through self-attention mechanisms in transformers. |

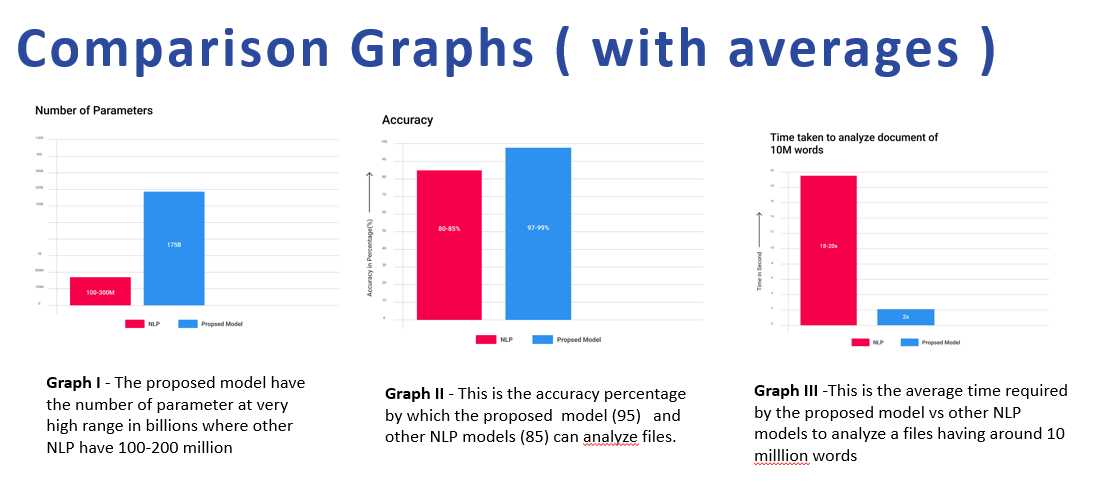
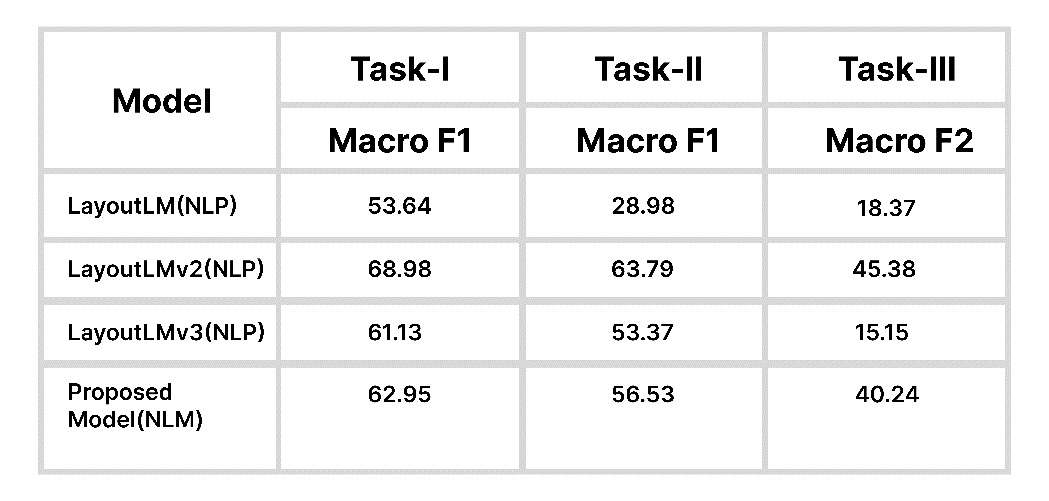
****  
  
In the forthcoming section, we present a comprehensive performance comparison. The evaluation is centered on critical metrics that include the number of parameters defining the   
model's complexity, the accuracy in processing and analysing textual data, and the time efficiency for document analysis. These metrics not only reflect the technical robustness of our proposed solution but also its practical superiority over conventional models in handling extensive linguistic data. Fig. 17 shows the comparison of NLP and the proposed model taking factors such as no. of parameters, accuracy and Time taken to analyse. This comparative analysis aims to demonstrate the significant advancements our model offers in the field of natural language processing.

Fig. 17. Performance comparison of NLP and LLM model

1. Reasoning

The tasks are designed to be similar to real applications. In Task 1 Single Template Learning, documents in the train and test sets are drawn from a single template. In Task 2 Mixed Template Learning, we increase the diversity of templates, but train and test sets for each document type are drawn from the same set of templates. In Task 3 Unseen Template Learning, the train and test sets are drawn from disjoint sets of templates to measure how well a model generalizes to unseen templates. Within each task, we compare the model performance with different number of training samples to understand the data efficiency for each approach. Discuss about table 3 and fig 16

 Table 3. Comparison of NLP and LLM model performance with our Proposed model

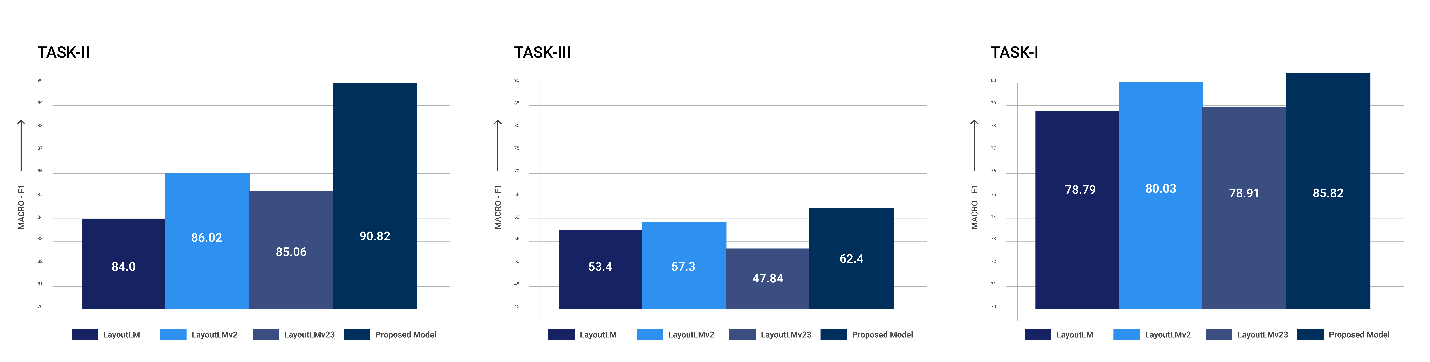


Fig. 16. Comparison of different models

**7.CONCLUSION AND FUTURE SCOPE**

The proposed AI-enabled interaction system offers promising avenues for future research and development. Blockchain optimization targets scalability, reduced costs, and enhanced access control. Advanced vectorization techniques promise better data representation, while multimodal AI models extend system capabilities. Ongoing efforts in security and privacy are critical. Scalability challenges and establishing robust feedback loops are pivotal for system growth and improvement. The research paper outlines an innovative system for file analysis within a Blockchain storage environment, featuring five essential layers for functionality and efficiency. This fusion of AI, Blockchain, and vector-based data analysis has transformative potential, providing secure, efficient, and intelligent solutions across various sectors.

**REFERENCES**

1. Zilong Wang Yichao Zhou w wei, C yu lee, S Tata. (2021). Visually Rich files Understanding (VRDU): A Benchmark for Information Extraction from Visually Rich filess. arXiv preprint arXiv:2102.06137.
2. Chenfei Wu, Shengming Yin, Weizhen Qi, Xiaodong Wang, Zecheng Tang Nan, Duan\* Visual ChatGPT: Talking, Drawing and Editing with Visual Foundation Models.
3. Robert Chew, John Bollenbacher,Michael Wenger (2021). LLM:assisted content analysis Using llm to support deductive coding.
4. Marzena Karpisnka, Mohit Iyyer (2023) LLM effectively leverage document-level context for literary translation but critical errors persists.
5. Yu-ting-Yang, Yan Ming Zhang, Xiao Long (2022) Casia-onDo: A New Database for Online Handwritten Document Analysis.
6. Le Chui, Yibeng Xu, Tengchao Lu (2021) Document Ai: benchmarks, Models and Applications.
7. Mu Cai, Zeyi Huang, Yuheng Li, Haohan Wang2, Yong Jae Lee1 (2023) Leveraging Large Language Models for Scalable Vector Graphics-Driven Image Understanding.
8. Yifan Zhang, Cheng Wei, Shangyou Wu, Zhenfting He, Wenhao Yu (2023). GeoGPT: Understanding and Processing Geospatial Tasks through An Autonomous GPT.
9. Cheonsu Jeong (2022). A Study on the Implementation of Generative AI Services Using an Enterprise Data-Based LLM Application Architecture
10. IAN ARAWJO, CHELSE SWOOPES, PRIYAN VAITHILINGAM (2023) ChainForge: A Visual Toolkit for Prompt Engineering and LLM Hypothesis Testing.
11. Chris Egersdoerfer, Di Zhang, Dong Dai (2021). Early Exploration of Using ChatGPT for Log-based Anomaly Detection on Parallel File Systems Logs.
12. Srikar Appalaraju, Bhavan Jasani, Bhargava Urala Kota, Yusheng Xie DocFormer: End-to-End Transformer for Document Understanding
13. Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova (2022) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding
14. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun (2021) Deep Residual Learning for Image Recognition
15. Zheng Huang, Kai Chen, Jianhua He, Xiang Bai, Dimosthenis Karatzas, Shijian Lu, C. V. Jawahar (2019) ICDAR2019 Competition on Scanned Receipt OCR and Information Extraction.
16. Minesh Mathew, Dimosthenis Karatzas, C.V. Jawahar (2021) DocVQA: A Dataset for VQA on Document Images
17. Rafa l Powalski, Lukasz Borchmann, Dawid Jurkiewicz (2021) Going Full-TILT Boogie on Document Understanding with Text-Image-Layout Transformer
18. Tomasz Stanisławek 1 , 2 , Filip Graliński 1 , 3 , Anna Wróblewska 2 , Dawid Lipiński 1 , Agnieszka Kaliska 1 , 3 , Paulina Rosalska 1 , Bartosz Topolski 1 , and Przemysław Biecek (2021) leister: Key Information Extraction Datasets Involving Long Documents with Complex Layouts.
19. Jiapeng Wang, Lianwen Jin, Kai Ding (2022) LiLT: A Simple yet Effective Language-Independent Layout Transformer for Structured Document Understanding.
20. Peng Zhang1∗ ,Yunlu Xu1∗ ,Zhanzhan Cheng21† ,Shiliang Pu1+ ,Jing Lu1 ,Liang Qiao1 ,Yi Niu1 , Fei Wu. (2021) TRIE: End-to-End Text Reading and Information Extraction for Document Understanding
21. Yang Xu1∗ , Yiheng Xu2 ∗ , Tengchao Lv2 ∗ , Lei Cui2 , Furu Wei2 , Guoxin Wang3 , Yijuan Lu3 , Dinei Florencio3 , Cha Zhang3 , Wanxiang Che1 , Min Zhang4 , Lidong Zhou2. (2022) LayoutLMv2: Multi-modal Pre-training for Visually-rich Document Understanding.
22. Chenchen Ding, Masao Utiyama, and Eiichiro Sumita. 2014. Document-level re-ranking with soft lexical and semantic features for statistical machine translation. In Proceedings of the 11th Conference of the Association for Machine Translation in the Americas: MT Researchers Track, pages 110–123, Vancouver, Canada. Association for Machine Translation in the Americas.
23. Wenxiang Jiao, Wenxuan Wang, Jen tse Huang, Xing Wang, and Zhaopeng Tu. 2023. Is ChatGPT A Good Translator? Yes with GPT-4 As The Engine.
24. Giorgos Vernikos, Brian Thompson, Prashant Mathur, and Marcello Federico. 2022. Embarrassingly Easy Document-level MT Metrics: How to Convert Any Pretrained Metric Into a Document-Level Metric. In Proceedings of the Seventh Conference on Machine Translation, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.
25. Katherine Thai, Marzena Karpinska, Kalpesh Krishna, Bill Ray, Moira Inghilleri, John Wieting, and Mohit Iyyer. 2022. Exploring document-level literary machine translation with parallel paragraphs from world literature. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, pages 9882–9902, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.