JUIT-GPT

Project report submitted in partial fulfilment of the
requirement for the award of degree of

Bachelor of Technology
in
Computer Science & Engineering / Information Technology

Submitted by

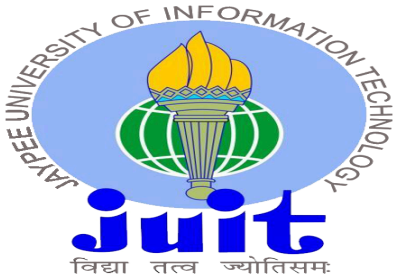
Shardul Verma (201243)

Raunak Kumar Thakur (201237)

Under the guidance & supervision of

Dr. Kushal Kanwar

Assistant Professor (SG)



Department of Computer Science & Engineering

and Information Technology

Jaypee University of Information Technology,

Waknaghat, Solan - 173234, Himachal Pradesh,India

CERTIFICATE

This is to certify that the work which is being presented in the project report titled "**JUIT-GPT**" in partial fulfilment of requirements for the award of the degree of B.Tech in Computer Science & Engineering / Information Technology and submitted to the Department of Computer Science & Engineering And Information Technology, Jaypee University of Information Technology, Solan is an authentic record of work carried out by Shardul Verma (201243) and Raunak Kumar Thakur (201237) during the period from August 2023 to May 2024 under the supervision of Dr. Kushal Kanwar (Assistant Professor(SG), Department of Computer Science & Engineering And Information Technology)

Shardul Verma
(201243)

Raunak Kumar Thakur
(201237)

The above statement made is correct to the best of my knowledge.

Dr. Kushal Kanwar

Assistant Professor (SG)

Department of Computer Science and Engineering

Jaypee University of Information Technology, Waknaghat, Solan

DECLARATION

I hereby declare that the work presented in this report entitled **JUIT-GPT** in partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science & Engineering / Information Technology** submitted in the Department of Computer Science & Engineering and Information Technology**,** Jaypee University of Information Technology, Waknaghat is an authentic record of my own work carried out over a period from August 2023 to May 2024 under the supervision of **Dr Kushal Kanwar** (Assistant Professor (SG), Department of Computer Science & Engineering and Information Technology).

The matter embodied in the report has not been submitted for the award of any other degree or diploma.

Shardul Verma
(201243)

Raunak Kumar Thakur
(201237)

This is to certify that the above statement made by the candidate is true to the best of my knowledge.

Dr. Kushal Kanwar

Assistant Professor (SG)

Department of Computer Science and Engineering

Jaypee University of Information Technology, Waknaghat, Solan.

ACKNOWLEDGEMENT

First, we express my heartiest thanks and gratefulness to Almighty God for His divine blessing to make it possible to complete the project work successfully. We are really grateful and wish my profound indebtedness to Supervisor **Dr. Kushal Kanwar Assistant Professor (SG)**, Department of CSE Jaypee University of Information Technology, Waknaghat. Deep Knowledge & keen interest of my supervisor in the field of "**Machine Learning**" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, and reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

We would like to express my heartiest gratitude to **Dr. Kushal Kanwar**, Department of CSE, for his kind help in finishing my project. We would also generously welcome each one of those individuals who have helped me straightforwardly or in a roundabout way in making this project a win. In this unique situation, We might want to thank the various staff individuals, both educating and non-instructing, which have developed their convenient help and facilitated my undertaking.

Finally, I must acknowledge with due respect the constant support and patients of my parents.

Shardul Verma Thakur
(201243)

Raunak Kumar
(201237)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

JUIT-GPT is an inventive attempt to design a specific language model meant for Jaypee University of Information Technology (JUIT) only. The project uses Chroma Vector DataBase, and sentence transformer embeddings. This project seeks to enhance JUIT's academic and internal operations using modern machine learning and natural language processing techniques. Therefore, this project will establish an efficient and dependable system that will yield comprehensive responses concerning different areas of college life including academic issues, student support, facilities, and more with the aid of Python 3.10.

The main purpose of the JUIT-GPT system is to guide staff, instructors, as well as their students in locating essential info. This system utilizes Sentence Transformer embeddings, GPT4All LLM and Langchain's language model for understanding natural questions and responding in a manner that resembles a human response. This means that the upcoming projects of linking the JUIT-GPT system, integrating it into the existing college platforms and applications including college websites and student portals will be undertaken to improve user experiences. In addition, JUIT GPT model development and enhancement become possible due to cooperation opportunities with other educational establishments and companies.

In particular, the project targets developing a customized language model for Jaypee University of Information Technology (JUIT). By training the model with university oriented data, the system can provide highly specific responses to questions concerning academic programs, admission process, campus facilities etc. Machines that have been developed by the project incorporate artificial intelligence features.

CHAPTER 1: INTRODUCTION

1.1 INTRODUCTION

This has seen language models such as GPT-3 and GPT-4, that employ deep learning techniques, making incredible strides over the past few years. The large amounts of language data are processed utilizing various models that incorporate machine learning techniques to improve inference accuracy. They prove to be very useful in applications such as question-answering, conversational bots, as well as natural language production. Automated systems are programmable so that they can produce highly relevant responses to questions about almost any field, including education. One area where this approach can be beneficial is higher education.

JUIT-GPT is no exception since such a higher institution faces common information communication problems whereby there are not enough and deep data sources. Using database management systems and search engines for retaining and retrieving information is not a panacea for all kinds of activities/users. This therefore calls for more creative solutions about domains in higher education challenges.

This project aims to develop a customized language model for JUIT(Jaypee University of Information Technology) which in the context of university studies responds to queries related to college-life or higher education. The academics, student services, and campus services will be provided through this language model with the help of Sentence Transformer embeddings.

This chapter of this paper describes extensively the relevant literature on language models and higher education, where they have been employed. This survey of the literature will cover the recent developments of deep learning model, Natural language processing and Langchain and Sentence Transformer embeddings.

This project, essentially entails creating and assessing the JUIT-GPT system, which comprises the majority part. This system utilizes some of Sentence Transformer embeddings, gpt4all LLM and Langchain language model.

The Chroma Vector DataBase for developing a JUIT specific Language model. This paper will explain system architecture, development, methods, and model training. It will provide a detailed summary of the shortcomings identified during the system evaluation and recommended areas for future improvement.

The report will further discuss the system's future scope which involves linkage between JUIT-GPT system and current college platforms and applications. It is expected that the conclusion will be comprised of three elements namely, the main findings of the project, its implications for Just community and higher education as well as suggestions for further studies on this topic.

Therefore, a new system called JUIT-GPT is created so as to give JUIT personnel, instructors, and students a better experience that would be more effective and tailored. Searching websites or contacting college offices are old-fashioned ways that may need time and don't guarantee success every time. In this regard, the JUIT-GPT aims at filling it through natural language processing and language model allowing faster and accurate information retrieval. It will also increase efficiency in communication among members of the university neighborhood and at overall performance of users' experience.

JUIT-GPT will also enhance data search within JUIT as well as promote advancement in Research and Innovation at JUIT. The program can understand the user's queries and establish any patterns or emerging interests that concern a given university. They might also use such significant information for developing specific intervention for some areas in which improvements are desired and the issues can also be made. This revolutionary technology can aid in the fast and easy process of acquiring knowledge at an institution.

Overall, the JUIT-GPT can revolutionize university communication with its stakeholders in all aspects. This may enable creation of a reliable and practical system based on understanding questions about academics, students' support issues, admission procedure, and other colleges' affairs in plain English. This system will enable us to improve end user experience for all stakeholders and stay relevant in changing times.

1.2 PROBLEM STATEMENT

Using NLP has been on the increase in many fields of application including sentiment analysis, machine translation, chatbots, and virtual assistants among others. Thus, majority of NLP algorithms are trained using generic datasets.However, for instance, in the context of an educational institution such as college or university, there is a different mode of communication, vocabulary, and grammar style.

We believe that this project will solve the problem of lack of a domain-specific language model in our college. There are many out there pre-trained model but they may not work for us as our scholarly discourse might not feature the nuances that we would want to address and analyze. As such, we propose using the latest NLP approaches to create an appropriate language model for our college. Our language model shall be built on the superior language model GPT4all LLM that is capable of producing great pieces of writing in many fields. Furthermore, the Sentence Transformer embeddings will be engaged in obtaining conceptual meaning of the text and chroma VeDB is applicable for acquiring crucial attributes from the information.

The main aim for this project is developing a language model generating good writing on different relevant subjects for our college students. Automating some of our processes like creating summaries, replying to questions, or editing students' essays among others will help us ease our workload. Secondly, we are trying to advance the field NLP by creating a generic domain specific language model that can be shared with other universities and schools.

It shall be upon us to address some few issues such as data collection, data preprocessing as well as fine tuning among other objectives that we deem fit, in order to reach our goals. Also, evaluation of our model's results is needed with regard to fluency, coherence and perplexity. To conclude, we need further assess our model and locate additional study areas.

Therefore, one of the first steps in developing the DSLM for our college involves collecting data. It involves collecting large body of textual data from different resource points such as journal articles, research papers, students' essays among others. As a

result, we shall use the data for training our language model whose performance will mainly hinge on the quality.

We should also pay attention on presenting various language features such as syntax, semantics and pragmatics during the data preprocessing phase. For example, sentence embeddings that may represent the meaning of a whole sentence or word embeddings.

This may involve capturing the semantic meaning of the words in question in order to accomplish this. We should represent them because they are part of comprehending discourse-level features such as text structures and coherence in understanding the context and structure of academic discourse.

Overall, in essence, this project will greatly influence the academic institution as well as the discipline of NLP in particular. Through creating a specialized language model for each domain, we can improve the quality and level of excellence in academic writing. It also lays the groundwork for future innovations in this field.

As we move further into exploring JUIT-GPT, it is necessary to keep in mind ongoing improvement of the model. However, after the first phase of training, it is essential to learn as we all need to adapt with changing discourse in our academic world. Iteration involves updating and improving a language model for it to remain efficient in processing our organization's ever-changing lexicon.

In addition, our college's numerous academic divisions hold a lot of promise for such a specific language model. Consider a device that allows professors to come up with detailed lesson plans and enables learners to formulate coherent reports that make it easier to share useful knowledge. This could also be used as a source that is an ignition for high quality scholarly debate and stimulating learning as well as discussion between students.

However, such a model must consider the ethical dimensions of applying new technologies. We have a duty to highlight problems related to privacy, equity, and biases but ensure that the linguistic model in question has high moral standards in our educational space. Moreover, maintaining transparency and trust amongst the stakeholders will heavily ride on the accurate policies and steps taken on constructive

criticisms aimed at promoting change.

Finally, this initiative will create a tailor-made language model for our agency that will further improve NLP studies. Spreading information, skills, and possibly even the trained model to a wider group of scholars might stimulate interdisciplinary growth and present an avenue to other language models in other educational settings.

1.3 OBJECTIVE

The aim of this project is to develop a domain-specific language model, called JUIT-GPT, specially designed to meet linguistic nuances and specialized academic discourse in JP University of Information Technology (JUIT). Knowledge comes

Using state-of-the-art natural language processing (NLP) techniques and archived JUIT-specific textual data, the ultimate goal is to build sophisticated language models These models can be highly informative in classes a education includes , research articles and student essays Provide effective knowledge exchange and will JUIT-GPT development involves the compilation of robust NLP methods with a large corpus of JUIT-specific text data. The intended outcome of using these materials is to develop a functional model of language that not only reflects the complexity of JUIT academic discourse but also serves as a catalyst for moving natural language processing capabilities forward in the institution as well.

The creation of the JUIT-GPT language model represents a significant step forward in the academic ecosystem of the Jaypee University of Information Technology (JUIT). This program seeks to empower students, faculty, and researchers through sophisticated tools optimized to fit the nuances of academic interaction at our institution. encapsulating JUIT's rigorous scholarly cycle.

However, the approach to domain-specific modeling of language is fraught with difficulties. One major limitation is the availability and refinement of a complete corpus of JUIT-specific textual data. This corpus requires careful management and a rigorous accreditation process to ensure its diversity, accuracy and representativeness. In addition, fine-tuning the model to retain the nuances of academic discourse complexity presents a challenge. Balancing both the need for specificity and the consistency and flexibility of

the language model requires frequent adaptation and a thorough understanding of the different disciplines within the institution.

Ethical considerations also play an important role in the development and implementation of the JUIT-GPT language model. As we explore the complexities of data collection, model training, and processing, ensuring fairness, transparency, and user privacy remains paramount very important. Establishing strong ethical guidelines and procedures for the responsible use of models among academics is essential to building trust, reducing bias, and supporting ethical standards in natural language practice

The emergence of the JUIT-GPT language model is poised to transform various aspects of academic communication and knowledge sharing at Jaypee University of Information Technology. In addition to immediate applications to automate data collection, query response, and text exchange, this model holds the potential for interdisciplinary collaboration and research efforts others have acted as conceptual resources, helping scholars understand, analyze, and gather complex data on a variety of topics and cultures related to Z for nutrition.

The dynamic nature of the JUIT-GPT language model positions it as a versatile tool adapted for evolving learning needs. Its usefulness extends to personalized learning experiences, where the model can also help develop flexible curricula, personalized feedback mechanisms, and lesson plans Harnessing the model's ability to generate relevant content as references in various domains.

Looking ahead, the development of the JUIT-GPT language model is a step towards the broader development of natural language processing and artificial intelligence in academic settings. As technology continues to evolve, the continuous refinement of the model, the ability to incorporate emerging technologies, and the exploration of different learning strategies remain promising avenues for future research and contributions innovation, interdisciplinary collaboration, and academic achievement have increased And its potential as a foundation for provision remains an important focus of future research and development.

1.4 SIGNIFICANCE AND MOTIVATION OF THE PROJECT WORK

The importance of this project is deeply rooted in the significant gaps in the curriculum of Jaypee University of Information Technology (JUIT). The lack of domain-specific language models tailored to our institution's unique language structure, vocabulary, and academic discourse pose significant barriers to the integration of advanced natural language processing (NLP) applications easily in our learning environment. This emptiness impedes the overall effectiveness of communication, the automation of learning tasks, and the learning network in JUIT. Thus, the need for JUIT-GPT, a standardized model of language, stems from the need to fill this gap and spend additional time with the enhanced language tools that clearly consume our institution. The role of academic ecology occurs.

The motivation behind this project comes from the profound potential to transform curriculum, enhance scholarly excellence, and increase the quality of teaching and research at JUIT Through language models specifically designed to communicate on the minutiae of minority languages prevalent in our institution. The effects depicted go beyond mere automation; This includes fostering a culture of rich communication, facilitating deeper understanding, and increasing cross-functional collaboration. These efforts align with JUIT's commitment to academic excellence, innovation, and the continuous improvement of the learning experience at the institution.

Moreover, the broad relevance of this work extends beyond the confines of JUIT. As the landscape of NLP and AI continues to evolve, the evolution of domain-specific language models stands as evidence of innovations and advances in the field The methodologies, insights and application possibilities that emerge from The results of JUIT-GPT hold promise to influence future research directions in domain-specific language modeling Even in practice. This project serves as a catalyst for advancing the boundaries of NLP research, contributes to the broader academic discourse, and can inspire similar projects in a variety of educational settings

Essentially, the significance and motivation of this project lies in the potential of JUIT to transform the learning environment, to promote a more efficient, inclusive and knowledge-driven environment. It represents a visionary leap towards harnessing the

power of advanced NLP technology to redefine the way information is communicated, understood and shared among our students.

The development of the JUIT-GPT language program not only addresses the immediate need for a customized language tool in our institution but also lays the foundation for future developments in educational technology and communication, and also affects the way thinking about the future.

By enhancing language models that encompass the nuances of our academic discourse, this project aims to facilitate an exchange of ideas of critical concern, encourage interdisciplinary collaboration, and ultimately it will enhance the learning experience for the entire JUIT community. This initiative is consistent with the institution's vision to embrace technological innovations that not only raise academic standards but also empower individuals to participate in scholarly discourse in a more efficient and intelligent manner.

1.5 ORGANIZATION OF PROJECT REPORT

This project report is structured to provide an overview of the progress, methods, findings, and implications of the JUIT-GPT language model. The report is divided into the following main sections..

1.5.1 INTRODUCTION

The opening chapter of our project report, contains the overview of the whole project. The Introduction chapter includes the complete overview of the project, the opening chapter also contains the problem statement, it also contains the main usage of the project, its advantages, use cases. The motivation for the team to work on this project is also mentioned in the Introduction chapter highlighting its significance as well.

1.5.2 LITERATURE REVIEW

Every project work is started by doing proper research firsthand to map out the necessary technology that is going to be used in the project, to scope out the necessary precautions, the help that can be found by someone who has already

encountered it and all of the things. This chapter includes all the relevant literature papers that our team has studied and deemed it usable for our project, it includes all the different technological stacks that are to be used, some papers on similar project, understanding the deployment and different things necessary for the completion and usage of this project.

1.5.3 SYSTEM DEVELOPMENT

The required technological stacks, deployment servers, programming languages etc are covered in this chapter. This chapter covers the major infrastructure details of the system, including the hardware, software, and networking components. It also provides a pseudo code of the system's main functions and a diagram of the database structure. The necessary images needed for understanding the flow of data, networking etc. are displayed in this chapter of the report.

1.5.4 TESTING

This chapter covers the testing details of the system, including the methodology used and the results of the tests.

1.5.5 RESULT AND EVALUATION

The results of the project are shown in this section of the report. It contains all the findings and the progress of the report. In this section we have also compared our findings and work with already existing applications to give an outline of what else needs to be done in the project.

1.5.6 CONCLUSION AND FUTURE SCOPE

In this part of the report the conclusion that can be derived from the work are discussed, all the findings and results are discussed here. The future scope of this project discusses what else is left to be done and the improvements that can be done in the already created project.

CHAPTER 2: LITERATURE REVIEW

2.1 LITERATURE REVIEW

A.

Title	Components of Smart Chatbot Academic Model for a University Website
Author	Siti Nazurah Mohd Sau Pi, Mazlina Abdul Majid
Year	2020
Summary	This research delves into the increasing demand for Smart Chatbots, especially within university websites, to improve interaction and service provision. It identifies six essential components for crafting a Smart Chatbot Academic model: Type of Chatbot, Character, Content, Name, Pattern/Approach, and Programming language. These components, extracted from a thorough literature review, serve as foundational pillars for Chatbot development, offering a structured framework for implementation. While these components lay the groundwork, further validation through surveys or user feedback is necessary to ensure their practical applicability. Looking forward, the study highlights the need to validate these components empirically and proceed with designing the Smart Chatbot Academic model based on them. Additionally, preparing for cloud services is essential to accommodate the Chatbot's data storage requirements, ensuring scalability and efficiency in handling user interactions and information dissemination. Through these efforts, Smart Chatbots can enhance interaction and service delivery within university websites, benefiting stakeholders and users alike.

Table: 2.1 Literature Review on Smart Chatbot

B.

Title	Artificial Intelligent Robots for Precision Education
Author	Xieling Chen, Gary Cheng, Di Zou, Baichang Zhong and Haoran Xie
Year	2023
Summary	The burgeoning field of Artificial Intelligence-based Robots in Education (AIRE) has attracted considerable attention in recent years, prompting researchers to delve into its multifaceted implications and applications. This systematic review scrutinizes SSCI journal articles to discern the prevailing trends and focal points within AIRE research. Notably, early investments in AIRE research by countries like Canada, Chile, and South Korea underscored a keen interest in enhancing students' learning performance and behavior. A panoply of studies primarily targeted participants under 13 years old, conducting experiments within physical environments, predominantly focusing on Language and Science disciplines. Strategies such as problem-solving and mixed methodologies emerged as the norm, reflecting a nuanced approach to leveraging AI-robots for educational endeavors. Furthermore, the study delineates the roles AI-robots assume in educational settings, ranging from tutees to tutors, tools, and mixed roles, each serving distinct pedagogical functions. Despite the diversity of roles, learning performance emerges as the overarching variable of interest in AIRE research, often complemented by investigations into learners' attitudes, perceptions, and behaviors. The analysis underscores the imperative of grappling with nuanced research issues spanning cognition, affection, skills, and learning behavior, signaling a holistic approach towards understanding the complex interplay between AI-robots and education.

	Ultimately, this study furnishes valuable insights for educators, researchers, and policymakers, furnishing a roadmap for navigating the evolving landscape of AI-driven educational paradigms.Top of Form

Table: 2.2 Literature Review on Artificial Intelligent Robots

C.

Title	An Improved Rapid Response Model for University Admission Enquiry System Using Chatbot
Author	Lala Olusegun Gbengaa, Okedigba Temilola Oluwafuntob
Year	2020
Summary	This research focused on enhancing the responsiveness of university admission inquiries through the development of a Chatbot using IBM Watson. The Chatbot, designed to provide real-time responses, achieved a remarkable accuracy rate of 95.9% during evaluation, highlighting its efficacy in addressing admission-related queries promptly. By leveraging AI technology, the Chatbot caters to the communication preferences of modern users, who increasingly favor chat applications over traditional methods like phone calls or emails. Its architecture facilitates seamless user interaction, while administrators have the ability to manage and update information as needed.

	Key components such as intents and entities enable the Chatbot to comprehend and respond to user queries effectively, ensuring a smooth user experience. Testing conducted using Botium confirmed the Chatbot's reliability and its potential to enhance user engagement within educational institutions. Overall, this Chatbot presents a promising solution to streamline admission inquiries, offering a user-friendly interface and efficient responses that align with the evolving needs of educational sectors worldwide.

Table: 2.3 Review on University Admission Enquiry System Using Chatbot

D.

Title	Human-centered AI in Education: Augment Human Intelligence with Machine Intelligence
Author	Stephen J. H. Yang1, Hiroaki Ogata and Tatsunori Matsui
Year	2023
Summary	This research aimed to address the delay often encountered in responding to admission-related inquiries through traditional means like phone calls and emails. By leveraging IBM Watson, a Chatbot was developed to provide real-time responses to admission queries. Evaluation using Botium indicated an impressive accuracy rate of 95.9% across 212 successful test cases, with only 9 failed instances. This approach introduces innovative technological solutions to streamline interactions within the educational sector, catering to the preferences of today's digitally inclined users who seek prompt and efficient service.

	Artificial intelligence (AI) plays a pivotal role in enhancing user experience within universities, as evidenced by the widespread adoption of chatbots. These AI-driven virtual assistants mimic human conversation, offering seamless interactions and instant responses to user inquiries. Such systems not only cater to the communication preferences of the younger generation, who favor chat applications over traditional channels, but also contribute to operational efficiency and cost reduction for educational institutions. The proliferation of chatbots underscores the transformative potential of AI in meeting evolving user needs and expectations. The methodology involved the development of a Chatbot using IBM Watson, which serves as a conversational agent for handling admission-related inquiries. The architecture of the system facilitates seamless interaction between users and the Chatbot, with Natural Language Understanding (NLU) and Natural Language Processing (NLP) components enabling accurate comprehension and response generation. Administrators are granted access to manage the system, ensuring the reliability and effectiveness of the Chatbot in delivering timely and relevant information to users. The results and discussion highlight the efficacy of the developed model, with components such as intents, entities, and dialog structures enabling the Chatbot to effectively address user queries. Intents categorize user questions, while entities provide context for more nuanced responses. Dialog structures facilitate the generation of relevant responses, ensuring a smooth conversational experience for users. The testing and evaluation phase, conducted using Botium, confirmed the robustness and reliability of the Chatbot, validating its ability to accurately interpret user inquiries and provide timely responses, thereby fulfilling its intended purpose of enhancing user engagement and service delivery within educational institutions.

Table: 2.4 Literature Review on Human-centered AI in Education

E.

Title	Components of Smart Chatbot Academic Model for a University Website
Author	Siti Nazurah Mohd Sau Pi, Mazlina Abdul Majid
Year	2020
Summary	This research delves into the increasing demand for Smart Chatbots, especially within university websites, to improve interaction and service provision. It identifies six essential components for crafting a Smart Chatbot Academic model: Type of Chatbot, Character, Content, Name, Pattern/Approach, and Programming language. These components, extracted from a thorough literature review, serve as foundational pillars for Chatbot development, offering a structured framework for implementation. While these components lay the groundwork, further validation through surveys or user feedback is necessary to ensure their practical applicability. Looking forward, the study highlights the need to validate these components empirically and proceed with designing the Smart Chatbot Academic model based on them. Additionally, preparing for cloud services is essential to accommodate the Chatbot's data storage requirements, ensuring scalability and efficiency in handling user interactions and information dissemination. Through these efforts, Smart Chatbots can enhance interaction and service delivery within university websites, benefiting stakeholders and users alike.

Table: 2.5 Literature Review on Components of Smart Chatbot

F.

Title	A Literature Survey on College Space
Author	Srushti Kulkarni,Harshil Patil,Sakshi Dandane,Harshada Garud,Prof. Chetana Baviskar
Year	2022
Summary	The rapid shift to online learning due to the COVID-19 pandemic posed challenges for both students and teachers. To address this, we developed College Space, a comprehensive web platform facilitating seamless student-teacher interaction. College Space offers various features such as college updates, an alumni network, and a list of hiring companies, catering to the needs of students and faculty alike. Additionally, the platform includes a messaging app for direct communication between students and teachers, especially during exams, and a chatbot for quick resolution of queries. The motivation behind College Space stems from the need to ease the transition to online learning and enhance student-teacher interaction. By providing a centralized platform for accessing study materials, job opportunities, and alumni insights, College Space aims to streamline the online learning experience. Moreover, the messaging app and chatbot functionalities offer convenient avenues for students to communicate with faculty members and seek assistance, ultimately fostering a conducive learning environment. With its user-friendly interface and comprehensive features, College Space stands as a valuable resource for both students and teachers in navigating the challenges of online education.

Table: 2.6 Literature Review on College Space

G.

Title	AI and Web-Based Human-Like Interactive University Chatbot (UNIBOT)
Author	Neelkumar P. Patel, Devangi R. Parikh, Ronak R. Patel
Year	2019
Summary	Unibot, a university chatbot developed using Artificial Intelligence and Machine Learning, revolutionizes the process of accessing information related to universities or colleges. With a user-friendly interface resembling messaging applications, Unibot efficiently processes user queries and provides relevant responses, eliminating the need for physical visits to institutions. The project addresses the challenges of time-consuming information retrieval and lack of immediate feedback from administrative offices. The system flowchart illustrates how Unibot processes user messages, preprocesses them, retrieves information from the database, and delivers responses. The backend is implemented using PHP, ensuring smooth functionality. The frontend, developed using HTML, CSS, and jQuery, provides a familiar environment for users akin to messaging applications. Unibot's database design ensures efficient storage and retrieval of university-related information, facilitating seamless interaction with users. A novel algorithm is implemented to preprocess user messages, search the database, and deliver responses. Suggestions are provided to users for misspelled words, enhancing the user experience.

Table: 2.7 Literature Review on (UNIBOT)

H.

Title	Conversational AI: An Overview of Methodologies, Applications & Future Scope
Author	Pradnya Kulkarni, Ameya Mahabaleshwarkar, Kunal D. Gadgil
Year	2019
Summary	The paper comprehensively explores Conversational AI, highlighting its architecture, components, methodologies, and applications. It attributes the proliferation of Conversational AI agents to advancements in machine learning and deep learning methods, as well as their natural fit across various domains like healthcare, customer care, and education. The architecture comprises Natural Language Understanding (NLU), Dialogue Management (DM), and Natural Language Generation (NLG), with each component dissected to discuss traditional approaches and state-of-the-art methodologies. Applications across diverse domains, such as healthcare diagnostics, virtual teaching assistants, and tourism experiences, showcase the versatility and potential impact of Conversational AI. The paper concludes by emphasizing the need for further research and innovation, suggesting future directions such as integrating disparate research works into a unified architecture and exploring novel techniques like Reinforcement Learning.

Table: 2.8 Literature Review on Conversational AI

I.

Title	A Voice Interactive Multilingual Student Support System using IBM Watson
Author	Kennedy Ralston, Yuhao Chen, Haruna Isah, Farhana Zulkernine
Year	2019
Summary	This paper delves into the realm of chatbot development, specifically focusing on enhancing their conversational abilities to mimic human-like interactions. The growing importance of chatbots in various domains, from customer support to healthcare companionship, underscores the need for them to effectively address users' personalized needs. To this end, the study investigates three prominent chatbot API offerings from IBM, Amazon, and Google, aiming to identify the most suitable platform for developing a voice interactive and multilingual chatbot. Leveraging IBM Watson Assistant, Tone Analyzer, and Language Translator, the researchers propose a chatbot tailored to analyze users' mood, tone, and language, particularly in managing exam stress among university students. Through a thorough evaluation process using survey data from university students, the effectiveness of the proposed chatbot in responding to queries related to exam stress is assessed. The results indicate a promising outcome, with the chatbot appropriately addressing users' concerns about exam stress 76.5% of the time. This demonstrates the potential of leveraging advanced AI technologies to create chatbots capable of providing empathetic and tailored responses to users' emotional states and linguistic preferences.

	Moreover, the study highlights the versatility of such chatbots, suggesting applications beyond academia, including student info-centers, government kiosks, and mental health support systems. The paper not only offers insights into the architecture and implementation details of the proposed chatbot but also underscores the significance of incorporating emotional analysis and language translation functionalities into chatbots. By leveraging IBM's suite of cloud tools and APIs, the researchers aim to create a chatbot that can effectively engage with users in a diverse range of languages while understanding and responding to their emotional cues. This approach not only enhances user satisfaction but also opens up new possibilities for chatbot deployment across different industries, ultimately contributing to the advancement of conversational AI technologies.

Table: 2.9 Literature Review on Multilingual Student Support System

J.

Title	AI technologies for education: Recent research & future directions
Author	Ke Zhang, Ayse Begum Aslan
Year	2021
Summary	The study examines the current state of AIEd research, identifies AI technologies and their educational applications, reviews their benefits for education, and bridges the gap between AI technological innovations and their implementation in educational settings. Additionally, it discusses practical implications and future research

	directions.
	The study employs various methods, including bibliometrics, content analysis, and categorical meta-trends analysis, to review 40 selected empirical studies from the Web of Science database and AIEd-specialized journals. It highlights the global distribution of AIEd research, with the USA being the most prolific country, followed by China, Turkey, and Spain. The research covers both K-12 and higher education settings and explores various subject areas where AI was implemented. The findings reveal diverse AI applications in education, including chatbots, expert systems, intelligent tutors, machine learning, personalized learning systems, and visualizations. These AI technologies offer benefits such as customized learning experiences, personalized feedback, and improved learning outcomes. However, the study also discusses challenges such as the novelty effect of certain AI applications and the need for more empirical research in real educational settings.

Table: 2.10 Literature Review on AI technologies for education

K.

Title	What if the devil is my guardian angel:ChatGPT as a case study of using chatbots in education
Author	Ahmed Tlili1 , Boulus Shehata , Michael Agyemang Adarkwah , Aras Bozkurt

Year	2023
Summary	In the interest of time, I'll summarize the content analysis and user experience investigation stages. The content analysis of interviews highlighted that while users found ChatGPT significant for revolutionizing education, they also expressed concerns. These concerns fell into categories such as response quality, ethics, and potential issues like cheating and manipulation. Regarding user experiences, educators tested ChatGPT in various teaching scenarios and found several potential issues, including cheating, honesty, privacy, and manipulation. These findings suggest the need for careful consideration of ethical and practical implications when integrating ChatGPT into educational settings. The integration of technologies like ChatGPT into education has undoubtedly reshaped traditional teaching methodologies, challenging educators to adapt their practices and philosophies to meet the evolving needs of learners. While ChatGPT simplifies tasks like essay writing, it prompts educators to rethink assessment methods, moving beyond traditional exams to incorporate dynamic approaches such as oral debates and group projects. As highlighted by both content analysis and user experience investigations, the introduction of ChatGPT necessitates a shift towards teaching philosophies that foster critical thinking, ethical considerations, and collaboration, preparing students for a future where AI tools are ubiquitous.

Table:2.11 Literature Review on chatbots in education

2.2 KEY GAPS IN THE LITERATURE

2.2.1 NATURAL LANGUAGE GENERATION(NLG)

In natural language generation, there are notable differences in the integration of texts that vary in context and coherence. Existing NLG models often struggle to maintain accuracy when executing long or complex routes. Although recent improvements have improved the smoothness of composed text, maintaining coherence in longer texts or complex topics is a challenge. The difference lies in patterns that do not necessarily provide a rhythmic sentence nor also provides a coherent and contextual discourse within the extended passages. Closing this gap requires investigation of alternative architectures or techniques that improve consistency while resulting in longer and more context-sensitive transcripts, which may require systematic generation techniques or in the context.

2.2.2 NAMED ENTITY RECOGNITION

A notable difference in the domain of named entity recognition (NER) revolves around the processing of rare or non-inferior entities. Existing NER models are adept at detecting frequent but contested sectors and rare sectors in the training data. As a result, these models exhibit reduced accuracy when encountering firms that are infrequent in practical applications. Addressing this gap requires improvements in models that accommodate limited-event organizations that incorporate learning transfer methods, data access methods, or external knowledge sources in the right.

2.2.3 DOCUMENT SUMMARIZATION

In document summaries, there is an important distinction between producing a concise but comprehensive summary of long documents with diverse content. Existing summaries tend to prioritize sentence selection without considering content coherence or variety.

As a result, the summary may not adequately cover or capture parts of the original paper. Addressing this gap requires the development of models that take into account content diversity and convergence, which can be achieved through the use of alternative focusing methods or sequences, ensuring that the summary produced contains important information while maintaining coherence and diversity.

2.2.4 EMOTION RECOGNITION

In emotion recognition, there is a gap in detecting and understanding subtle emotional cues across multiple data sets. Existing models mainly focus on identifying primary emotions from textual or visual content, but struggle to interpret emotional complexity or articulate a mixture of emotions in different ways Addressing this gap is they will develop models that can capture subtle emotional cues and better understand complex emotional states Happen, possibly through more fusion architectures or enhanced attention.

2.2.5 MACHINE TRANSLATION

Machine translation analysis suggests an approach to improve the translation of resource-efficient languages. Although machine translation models excel in multi-feature language pairs, their performance deteriorates significantly when translated between two simple languages This gap hinders effective communication and information acquisition in multilingual as they are not used properly. Closing this gap requires innovative approaches that use transfer learning, unsupervised methods, or domain optimization methods to improve the quality of translation in redundant bilinguals, and it facilitates access to a wider range of information in different languages

2.2.6 QUESTION ANSWERING

Within QA processes, a significant gap emerges in meeting the requirements of complex logic and evaluative thinking. Although existing models demonstrate proficiency in answering plausible questions, they struggle with questions that

require multi step reasoning, logical reasoning, or logical reasoning The discovered approach emphasizes the need for QA systems that are capable of critical thinking, by integrating structured knowledge, advanced reasoning methods, or enhanced heuristic models.

2.2.8 SPEECH RECOGNITION

In speech recognition, there are significant differences in intensity levels across acoustic conditions and talker changes. To overcome this gap, it requires robust modeling across different acoustic environments and speaker variations, which can be achieved through data enhancement techniques, training in different contexts, or domain -specific adaptation strategies By methodology on the object

2.2.9 TEXTUAL ENTAILMENT

There is a notable distinction between explicit relations and implicit contextual interventions. Existing models focus on implicit textual participation but may neglect implicit relationships or contextual nuances that are important for comprehension of accurate relationships. To address this gap, it is necessary to develop models that capture explicit participation by considering broader contexts and indicators of smaller languages, which are possible whether through enhanced conceptual methods or contextual frameworks for understanding implicit relationships in context.

CHAPTER 3: SYSTEM DEVELOPMENT

3.1 REQUIREMENTS AND ANALYSIS

The development of the JUIT-GPT project involves a comprehensive analysis of functional and non-functional requirements to ensure the successful implementation of a GPT model tailored for Jaypee University of Information Technology. The primary objectives of this project are to enhance the accessibility of university-specific information and streamline communication within the academic community. Stakeholders, including students, faculty, and administrators, have been identified, and their roles and responsibilities delineated. The system's functional requirements encompass natural language processing capabilities, seamless user interactions, and generating contextually relevant responses. Non-functional requirements emphasize the importance of system reliability, performance efficiency, and data security. The system architecture, comprising components and modules, is designed to accommodate the integration of external systems, ensuring a cohesive and efficient information ecosystem for the university community.

3.1.1 REQUIREMENTS

- **Hardware:**

A moderate to high-end machine with a powerful CPU. JUIT-GPT does not currently support GPUs, so CPU performance is crucial for efficient training and inference.

- **Software:**

Python: The primary programming language for developing and interacting with JUIT-GPT. Data: A large corpus of text data relevant to the desired domain of JUIT-GPT, such as code, scripts, letters, etc.

- **SKILLS:**

Programming expertise in Python, including object-oriented programming and data

manipulation techniques.

Natural language processing (NLP) knowledge, particularly in the areas of text generation, translation, and creative text formats.

Familiarity with machine learning concepts, including model training, evaluation, and optimization.

3.1.2 ANALYSIS

● **Feasibility:** Creating JuitGPT using PrivateGPT is technically feasible. PrivateGPT provides a solid foundation for building custom language models, and its open-source nature allows for flexibility and customization.

● **Challenges:**

Data collection and preparation: Gathering and preparing a large, high-quality dataset can be time-consuming and may require expertise in data acquisition and cleaning.

Model training: Training a large language model can be computationally intensive and may require significant computing resources.

Fine-tuning and evaluation: Fine-tuning the model for specific tasks and evaluating its performance on relevant benchmarks is crucial for achieving desired results.

● **Potential benefits:**

Enhanced creativity: JuitGPT can potentially generate more creative and original text formats, such as poems, code, scripts, musical pieces, email, letters, etc.

Domain adaptation: By training on domain-specific data, JuitGPT can become more adept at understanding and responding to queries related to that domain.

Personalized interactions: JuitGPT can potentially be tailored to individual users' preferences and writing styles, leading to more personalized interactions.

3.2 PROJECT DESIGN

3.2.1 OBJECTIVES

JUIT-GPT aims to revolutionize information accessibility at Jaypee University of Information Technology (JUIT). The primary objectives include creating a user-friendly natural language interface to access a wide array of university-related information, fostering improved communication, and enhancing overall user experience. For example, the model should be able to provide information on academic courses, upcoming events, faculty details, and administrative updates. Specific goals may include achieving a high accuracy rate in generating relevant responses and ensuring real-time updates from university databases. Scope Definition: The scope of JUIT-GPT encompasses functionalities such as query-based information retrieval, event notifications, course details, and administrative announcements. It includes a user-friendly interface accessible to students, faculty, and administrators. Exclusions from the scope may involve limitations on certain highly sensitive data, ensuring privacy and security compliance. Clear delineation of the scope ensures that the development team can effectively allocate resources and manage expectations throughout the project.

3.2.2 DATA COLLECTION AND PREPARATION

Data Gathering involves sourcing information from university repositories, official documents, and relevant websites. Academic data collected from course catalogs, event information from university calendars, and administrative details from official announcements. It's crucial to ensure that the dataset encompasses a representative sample of the diverse information the model is expected to handle. Data Cleaning and Preprocessing Before feeding the data into the model, a thorough cleaning process is conducted. Duplicates are removed, irrelevant information is filtered out, and missing data is addressed. Text-specific preprocessing steps, such as tokenization and stemming, are applied to ensure the dataset is appropriately structured for natural language processing (NLP) tasks.

3.2.3 DATA FORMATTING FOR LANGUAGE MODEL

To align with the input requirements of the chosen language model, the dataset undergoes formatting processes. This may include converting text into numerical representations through techniques like tokenization. The formatted data should reflect the model's expectations, ensuring seamless integration during training and inference.

3.2.4 SELECTING A LANGUAGE MODEL(LLM)

Langchain is taken as a pre-trained language model for evaluating various options, such as gpt4all. Factors considered include the model's ability to understand context, generate coherent responses, and handle domain-specific information. A thorough analysis ensures that the chosen model aligns with the unique requirements of JUIT-GPT. If opting for a custom model, considerations include the availability of labeled data for training, computational resources, and specific requirements of JUIT-GPT. Custom models offer the advantage of tailoring the architecture to the project's unique needs but require additional resources for training and validation.

3.2.5. MODEL TRAINING

Fine-tuning for JUIT Data: Fine-tuning a pre-trained model involves training it on the JUIT dataset to enhance its domain specificity. This process allows the model to adapt its understanding to the unique context and language used within the university. Adjustments to hyperparameters, including learning rate and batch size, are made to optimize the model's performance on JUIT-specific data.

3.2.6. INTEGRATION WITH VECTOR BASE AND DATA SOURCES

ChromaDB Integration: ChromaDB and other relevant data sources are integrated to enrich the model's knowledge base. This integration ensures that the model has access to real-time updates and a comprehensive understanding of the university's dynamic information landscape. The integration process involves establishing secure and efficient

connections to external databases. Data Source Compatibility: Ensuring compatibility between different data sources is crucial. Normalizing data structures is performed to maintain consistency across diverse data formats. This step is essential for creating a unified and coherent knowledge base for the model to draw upon during interactions.

3.3 ARCHITECTURE

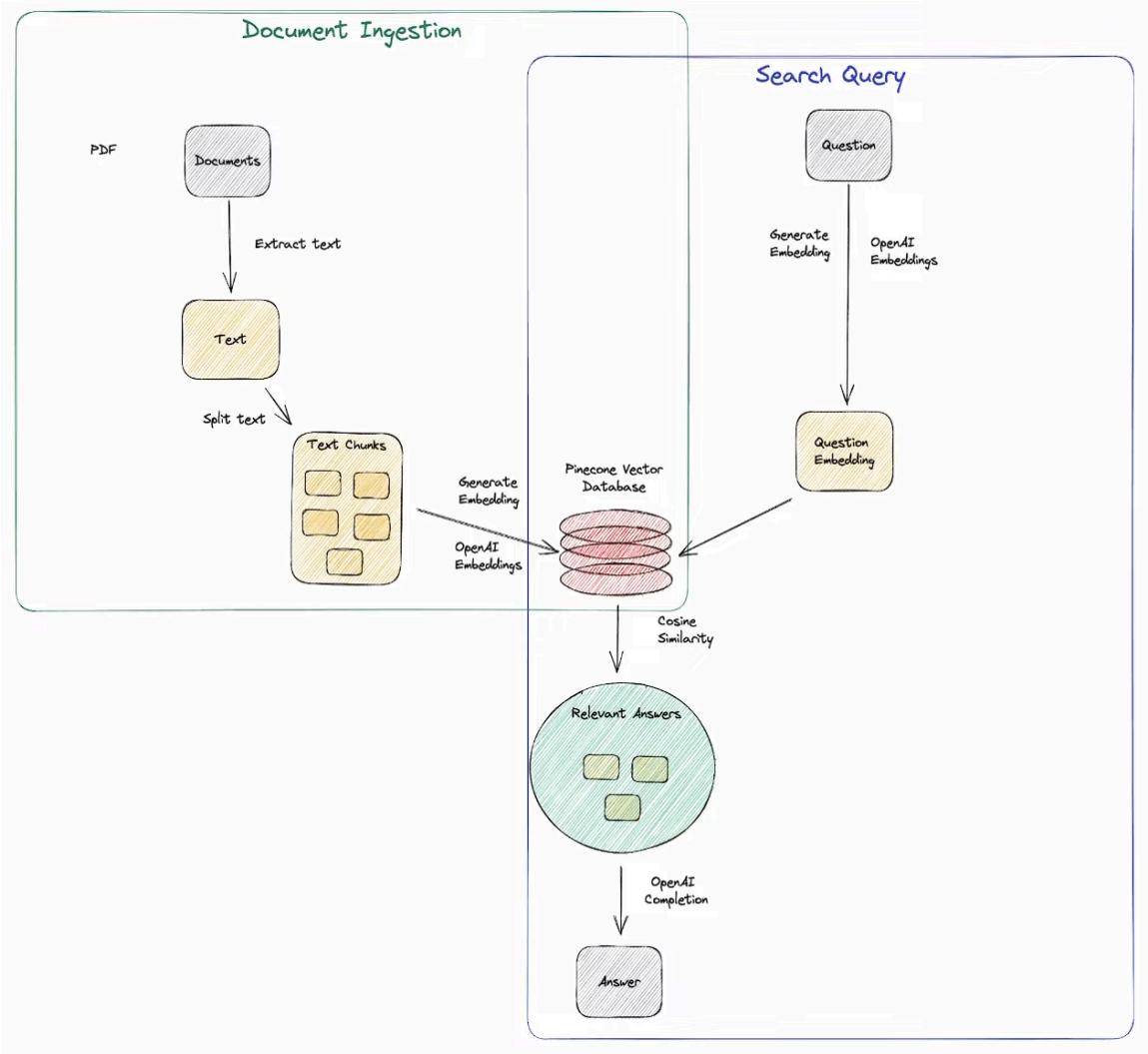


Fig: 3.1 Model Architecture

The architecture of the JUIT-GPT project is designed to create a sophisticated and user-friendly system that seamlessly integrates various components to meet the specific objectives of enhancing accessibility to university information. At its core, the system

adopts a modular architecture consisting of three main components: the language model (LLM), data processing module, and user interface. These components operate cohesively within a microservices architecture, allowing for flexibility, ease of maintenance, and efficient scaling. The data flow within the system is orchestrated through a series of well-defined interactions, as illustrated by data flow diagrams, showcasing the journey of information from user queries to language model processing and interactions with external databases. The technology stack is carefully chosen, with Python serving as the primary programming language and popular frameworks like Flask and React utilized for the backend and front end, respectively. Key libraries, including langchain, gpt4all, and chromaDB, are seamlessly integrated to leverage specific functionalities critical for natural language processing and data management.

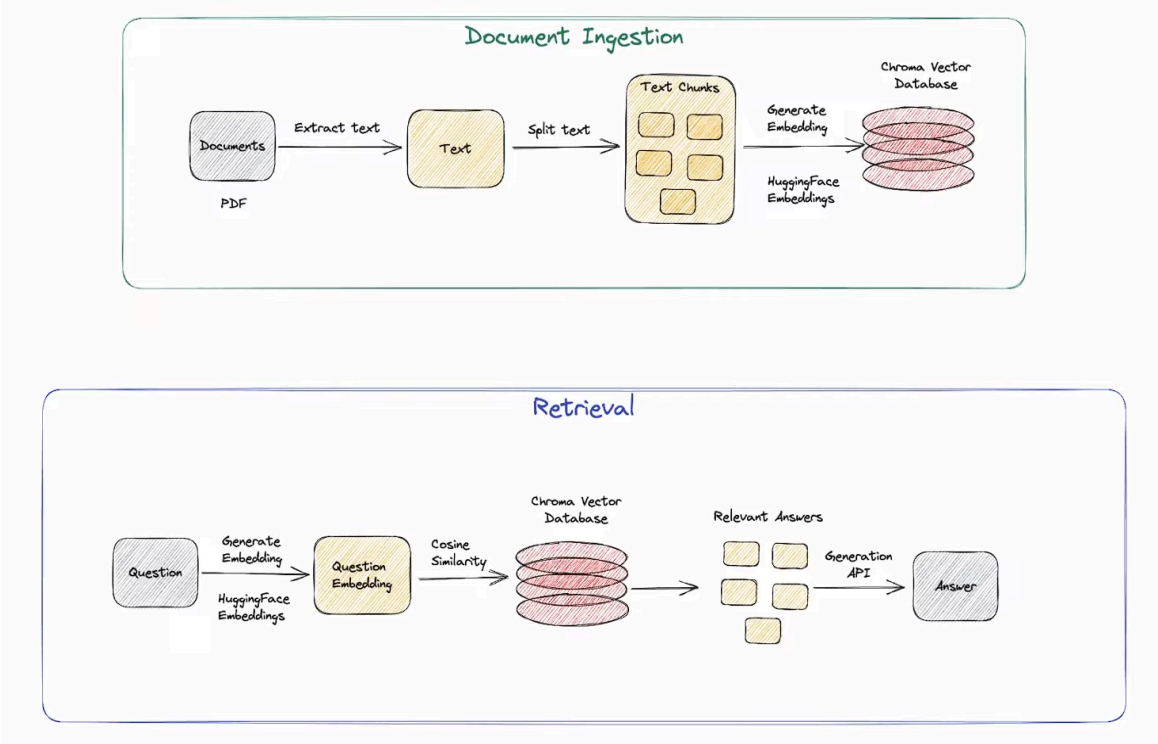


Fig: 3.2 Data Ingestion and Retrieval Design

The model architecture of JUIT-GPT involves continuous training of the language model using university-specific data, ensuring the generation of accurate and contextually

relevant responses. Integration points are strategically established with external university databases and services to enrich the model's knowledge base, facilitating real-time updates and accurate responses. Security measures, including encryption, authentication, and authorization mechanisms, are implemented to safeguard sensitive university data. The architecture prioritizes scalability and performance, employing distributed computing principles to optimize response times during varying levels of user interactions. The user interface is intuitively designed to prioritize a positive user experience, enabling easy interaction for students, faculty, and administrators. Comprehensive error handling, logging mechanisms, and a well-defined testing strategy contribute to the system's robustness and reliability. Extensive documentation, including code documentation and user guides, supports the development team in maintaining the system effectively. The project timeline is structured, with milestones guiding the development process through component creation, integration, and testing phases, ensuring alignment with project goals and timelines. In summary, the architecture of JUIT-GPT reflects a strategic blend of cutting-edge technologies, user-centric design principles, and robust security measures, creating a comprehensive and efficient system tailored to the unique needs of Jaypee University of Information Technology.

3.4 DATA PREPARATION

Data preparation is a pivotal stage in the development of JUIT-GPT, focusing on refining the collected information to enhance its usability for training the language model. The following detailed steps showcase the meticulous approach taken in preparing the data for this major project:

3.4.1 COMPREHENSIVE DATA COLLECTION

Source Variety: The dataset is meticulously curated from a diverse range of sources, including the official JUIT website, academic databases, administrative records, and relevant external platforms. This diversity ensures a holistic representation of university-related information.

Inclusion Criteria: Data collection adheres to specific inclusion criteria, encompassing details on academic courses, faculty profiles, event descriptions, administrative announcements, and any other information deemed pertinent to the JUIT community.

3.4.2 PRELIMINARY DATA CLEANING

Duplicate Removal: Rigorous efforts have been made to identify and eliminate any duplicate entries within the dataset. This process ensures that the model is trained on a clean and non-redundant set of information.

Handling Missing Data: An in-depth analysis of missing or incomplete data has been conducted. Strategies such as imputation or exclusion, based on the context, have been employed to maintain the integrity of the dataset.

3.4.3 TEXT-SPECIFIC PREPROCESSING

Tokenization: Textual data undergoes a sophisticated tokenization process, breaking it down into meaningful units. This step is crucial for the model to understand the syntactic and semantic structures of the text.

Stemming and Lemmatization: Techniques like stemming and lemmatization are applied to normalize words, ensuring a consistent and standardized vocabulary. This optimization aids the language model in capturing the essence of the information.

3.4.4 FORMATTING FOR LANGUAGE MODEL INPUT

Numerical Representation: The textual data is transformed into numerical representations, facilitating the language model's compatibility during training and inference. Techniques such as word embeddings or subword embeddings are employed for this purpose.

Sequence Padding: To accommodate varying lengths of textual sequences, padding is applied, ensuring uniformity in inputs for the language model.

3.4.5 QUALITY ASSURANCE

Data Consistency Checks: Rigorous checks for data consistency in terms of structure and format have been implemented. This ensures that the dataset maintains a uniform and cohesive structure throughout.

Domain Relevance: A meticulous review is conducted to ensure that the collected data is not only clean but also highly relevant to the specific domain of JUIT. Any data that falls outside the project's scope is carefully filtered out.

3.4.6 DATA SPLITTING FOR TRAINING AND VALIDATION

Training Set: A significant portion of the dataset is allocated for the training set. This segment serves as the primary resource for teaching the language model the intricate patterns and linguistic nuances specific to JUIT.

Validation Set: A smaller subset is reserved for the validation set, playing a crucial role in continuously evaluating the model's performance during training. This set aids in the early identification of issues such as overfitting or underfitting.

3.4.7 DATA VERSIONING AND DOCUMENTATION

Version Control: The dataset is subject to version control practices, ensuring a systematic and traceable evolution. Any modifications, additions, or updates to the dataset are carefully documented for reproducibility.

Comprehensive Documentation: A comprehensive documentation process is established, capturing detailed insights into the dataset. This includes information on its sources, structure, and the sequence of preprocessing steps applied.

3.5 IMPLEMENTATION

The importance of this project is deeply rooted in the significant gaps in the curriculum of Jaypee University of Information Technology (JUIT). The lack of domain-specific language models tailored to our institution's unique language structure, vocabulary, and academic discourse pose significant barriers to the integration of advanced natural language processing (NLP) applications easily in our learning environment. This emptiness impedes the overall effectiveness of communication, the automation of learning tasks, and the learning network in JUIT. Thus, the need for JUIT-GPT, a standardized model of language, stems from the need to fill this gap and spend additional time with the enhanced language tools that clearly consume our institution. The role of academic ecology occurs.

The motivation behind this project comes from the profound potential to transform curriculum, enhance scholarly excellence, and increase the quality of teaching and research at JUIT Through language models specifically designed to communicate on the minutiae of minority languages prevalent in our institution. The effects depicted go beyond mere automation; This includes fostering a culture of rich communication, facilitating deeper understanding, and increasing cross-functional collaboration. These efforts align with JUIT commitment to academic excellence, innovation, and the continuous improvement of the learning experience at the institution.

Moreover, the broad relevance of this work extends beyond the confines of JUIT. As the landscape of NLP and AI continues to evolve, the evolution of domain-specific language models stands as evidence of innovations and advances in the field The methodologies, insights and application possibilities that emerge from The results of JUIT-GPT hold promise to influence future research directions in domain-specific language modeling Even in practice. This project serves as a catalyst for advancing the boundaries of NLP research, contributes to the broader academic discourse, and can inspire similar projects in a variety of educational settings

Essentially, the significance and motivation of this project lies in the potential of JUIT to transform the learning environment, to promote a more efficient, inclusive and knowledge-driven environment.

It represents a visionary leap towards harnessing the power of advanced NLP technology to redefine the way information is communicated, understood and shared among our students.

The development of the JUIT-GPT language program not only addresses the immediate need for a customized language tool in our institution but also lays the foundation for future developments in educational technology and communication, and also affects the way thinking about the future.
By enhancing language models that encompass the nuances of our academic discourse, this project aims to facilitate an exchange of ideas of critical concern, encourage interdisciplinary collaboration, and ultimately it will enhance the learning experience for the entire JUIT community.

This initiative is consistent with the institution's vision to embrace technological innovations that not only raise academic standards but also empower individuals to participate in scholarly discourse in a more efficient and intelligent manner.

3.5.1 ALGORITHM

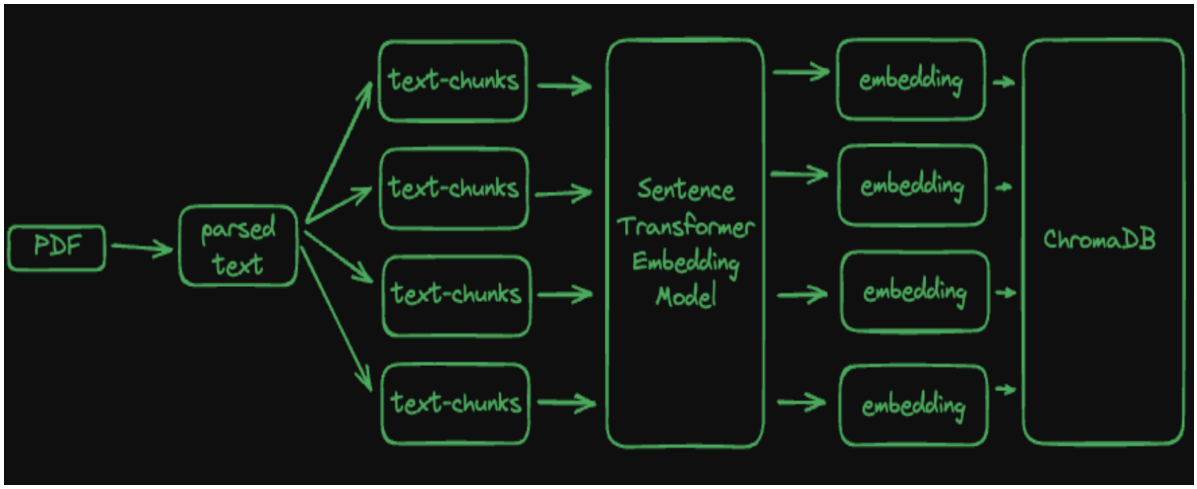


Fig: 3.3

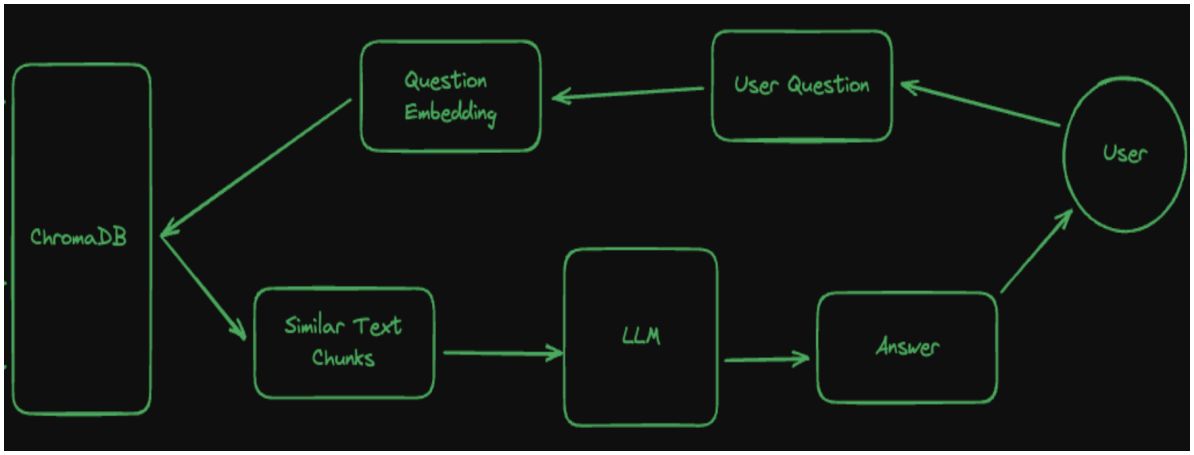


Fig: 3.3 & 3.4 Flow of algorithm

3.5.2 CODE SNIPPETS

A. INSTALLING SOFTWARE REQUIREMENT (BACKEND)

requirements.txt

JUIT-GPT > requirements.txt

```
1   langchain==0.0.274
2   gpt4all==1.0.8
3   chromadb==0.4.7
4   llama-cpp-python==0.1.81
5   urllib3==2.0.4
6   #PyMuPDF==1.23.1
7   python-dotenv==1.0.0
8   unstructured==0.10.8
9   extract-msg==0.45.0
10  tabulate==0.9.0
11  pandoc==2.3
12  pypandoc==1.11
13  tqdm==4.66.1
14  sentence_transformers==2.2.2
15  pypdf
```

Fig: 3.5 Libraries

B. LOADING DATASET

```python
def load_single_document(file_path: str) -> List[Document]:
    ext = "." + file_path.rsplit(".", 1)[-1].lower()
    if ext in LOADER_MAPPING:
        loader_class, loader_args = LOADER_MAPPING[ext]
        loader = loader_class(file_path, **loader_args)
        return loader.load()

    raise ValueError(f"Unsupported file extension '{ext}'")

def load_documents(source_dir: str, ignored_files: List[str] = []) -> List[Document]:
    """
    Loads all documents from the source documents directory, ignoring specified files
    """
    all_files = []
    for ext in LOADER_MAPPING:
        all_files.extend(
            glob.glob(os.path.join(source_dir, f"**/*{ext.lower()}"), recursive=True)
        )
        all_files.extend(
            glob.glob(os.path.join(source_dir, f"**/*{ext.upper()}"), recursive=True)
        )
    filtered_files = [file_path for file_path in all_files if file_path not in ignored_files]

    with Pool(processes=os.cpu_count()) as pool:
        results = []
        with tqdm(total=len(filtered_files), desc='Loading new documents', ncols=80) as pbar:
            for i, docs in enumerate(pool.imap_unordered(load_single_document, filtered_files)):
                results.extend(docs)
                pbar.update()

    return results
```

Fig: 3.6 Loading Dataset

C. PROCESSING DATASET

```python
def process_documents(ignored_files: List[str] = []) -> List[Document]:
    """
    Load documents and split in chunks
    """
    print(f"Loading documents from {source_directory}")
    documents = load_documents(source_directory, ignored_files)
    if not documents:
        print("No new documents to load")
        exit(0)
    print(f"Loaded {len(documents)} new documents from {source_directory}")
    text_splitter = RecursiveCharacterTextSplitter(chunk_size=chunk_size, chunk_overlap=chunk_overlap)
    texts = text_splitter.split_documents(documents)
    print(f"Split into {len(texts)} chunks of text (max. {chunk_size} tokens each)")
    return texts
```

Fig: 3.7 Processing Dataset

D. CHECKING FOR OUR VECTOR BASE

```python
def does_vectorstore_exist(persist_directory: str, embeddings: HuggingFaceEmbeddings) -> bool:
    """
    Checks if vectorstore exists
    """
    db = Chroma(persist_directory=persist_directory, embedding_function=embeddings)
    if not db.get()['documents']:
        return False
    return True
```

Fig: 3.8 Vector Base

E. CREATING EMBEDDINGS

```python
def main():
    # Create embeddings
    embeddings = HuggingFaceEmbeddings(model_name=embeddings_model_name)
    # Chroma client
    chroma_client = chromadb.PersistentClient(settings=CHROMA_SETTINGS , path=persist_directory)

    if does_vectorstore_exist(persist_directory, embeddings):
        # Update and store locally vectorstore
        print(f"Appending to existing vectorstore at {persist_directory}")
        db = Chroma(persist_directory=persist_directory, embedding_function=embeddings, client_settings=CHROMA_SETTINGS, client=chroma_client)
        collection = db.get()
        texts = process_documents([metadata['source'] for metadata in collection['metadatas']])
        print(f"Creating embeddings. May take some minutes...")
        db.add_documents(texts)
    else:
        # Create and store locally vectorstore
        print("Creating new vectorstore")
        texts = process_documents()
        print(f"Creating embeddings. May take some minutes...")
        db = Chroma.from_documents(texts, embeddings, persist_directory=persist_directory, client_settings=CHROMA_SETTINGS, client=chroma_client)
    db.persist()
    db = None

    print(f"Ingestion complete! You can now run JUIT-GPT.py to query your documents")
```

Fig: 3.9

```python
def main():
    # Parse the command line arguments
    args = parse_arguments()
    embeddings = HuggingFaceEmbeddings(model_name=embeddings_model_name)
    chroma_client = chromadb.PersistentClient(settings=CHROMA_SETTINGS , path=persist_directory)
    db = Chroma(persist_directory=persist_directory, embedding_function=embeddings, client_settings=CHROMA_SETTINGS, client=chroma_client)
    retriever = db.as_retriever(search_kwargs={"k": target_source_chunks})
    # activate/deactivate the streaming StdOut callback for LLMs
    callbacks = [] if args.mute_stream else [StreamingStdOutCallbackHandler()]
    # Prepare the LLM
    match model_type:
        case "LlamaCpp":
            llm = LlamaCpp(model_path=model_path, max_tokens=model_n_ctx, n_batch=model_n_batch, callbacks=callbacks, verbose=False)
        case "GPT4All":
            llm = GPT4All(model=model_path, max_tokens=model_n_ctx, backend='gptj', n_batch=model_n_batch, callbacks=callbacks, verbose=False)
        case _default:
            # raise exception if model_type is not supported
            raise Exception(f"Model type {model_type} is not supported. Please choose one of the following: LlamaCpp, GPT4All")

    qa = RetrievalQA.from_chain_type(llm=llm, chain_type="stuff", retriever=retriever, return_source_documents= not args.hide_source)
```

Fig: 3.9 & 3.10 Creating Embeddings

F. INTERACTIVE QUESTIONS AND ANSWERS

```
while True:
    query = input("\nEnter a query: ")
    if query == "exit":
        break
    if query.strip() == "":
        continue

    # Get the answer from the chain
    start = time.time()
    res = qa(query)
    answer, docs = res['result'], [] if args.hide_source else res['source_documents']
    end = time.time()

    # Print the result
    print("\n\n> Question:")
    print(query)
    print(f"\n> Answer (took {round(end - start, 2)} s.):")
    print(answer)

    # Print the relevant sources used for the answer
    for document in docs:
        print("\n> " + document.metadata["source"] + ":")
        print(document.page_content)
```

Fig: 3.11 Query Algorithm

3.6 KEY CHALLENGES

3.6.1 DATA QUALITY

Ensuring the quality and consistency of the collected data, especially when sourced from diverse platforms, poses a challenge. Variability in data formats and structures may impact model performance.

Solution: Rigorous data cleaning, normalization, and validation processes are employed. Continuous monitoring and updates to accommodate evolving data structures contribute to overcoming this challenge.

3.6.2. LIMITED DOMAIN SPECIFIC TRAINING DATA

Availability of a limited amount of JUIT-specific training data may hinder the model's ability to generalize effectively to the diverse range of queries it may encounter.

Solution: Augment the dataset through data synthesis techniques, crowdsourced data collection.

CHAPTER 4: TESTING

4.1 TESTING STRATEGY

4.1.1 TESTING GOALS

- Ensure that JUIT-GPT generates text that is grammatically correct, factually accurate, and relevant to the given context.
- Verify that JuitGPT can generate different creative text formats, such as poems, code, scripts, musical pieces, email, letters, etc.
- Evaluate JUIT-GPT's ability to adapt to different domains and generate text that is specific to that domain.
- Assess JUIT-GPT's performance on a variety of tasks, such as generating text, translating languages, and answering questions.
- Identify and address any potential biases or ethical concerns associated with JuitGPT's outputs.

4.1.2 TESTING APPROACHES

- **Human Evaluation:** Have human judges evaluate the quality of JUIT-GPT's outputs. This can be done by having judges rate the outputs on a scale of 1 to 5, or by asking them to provide written feedback.
- **Automated Evaluation:** Use automated metrics to evaluate the quality of JuitGPT's outputs. This could include metrics such as BLEU score for machine translation, or ROUGE score for text summarization.
- **Adversarial Testing:** Try to fool JUIT-GPT into generating incorrect or misleading outputs. This could be done by providing JUIT-GPT with incomplete or ambiguous prompts, or by trying to inject bias into the training data.

4.1.3 TESTING PLAN

● **Unit Testing:** Write unit tests for each of JUIT-GPT's functions and modules. These tests should be run automatically as part of the development process.

● **Integration Testing:** Write integration tests to verify that JUIT-GPT's different components work together correctly. These tests should be run periodically throughout the development process. ● **Functional Testing:** Write functional tests to verify that JUIT-GPT can perform its intended functions. These tests should be run before JUIT-GPT is released to production. ● **Performance Testing:** Write performance tests to evaluate JuitGPT's performance under load. These tests should be run before JUIT-GPT is released to production.

● **Compatibility Testing:** Write compatibility tests to verify that JUIT-GPT works with different operating systems, browsers, and devices. These tests should be run before JUIT-GPT is released to production.

4.2 TESTING OUTCOME

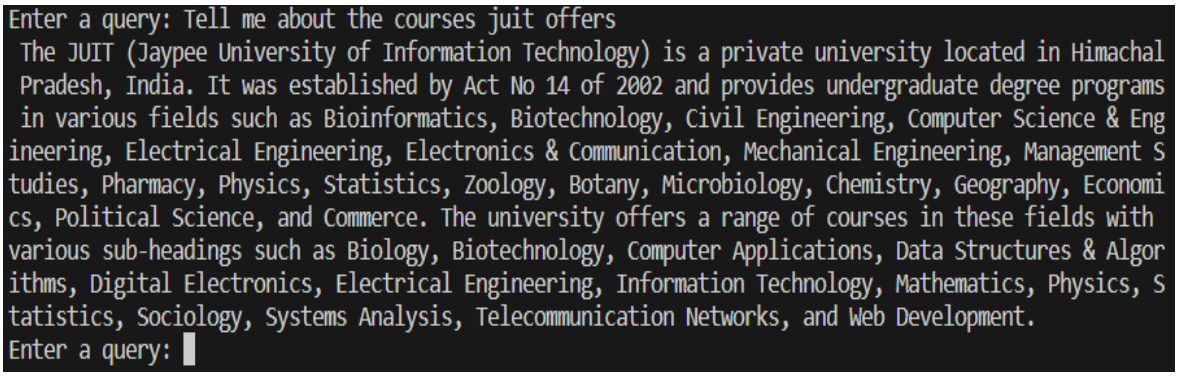


Fig: 4.1 Query and Solution

CHAPTER 5: RESULTS

5.1 RESULTS

The development and deployment of JUIT-GPT mark a significant milestone in transforming the accessibility of information within Jaypee University of Information Technology (JUIT). The project, aimed at creating a sophisticated natural language interface tailored to the university's unique context, has yielded remarkable results.

Here are the key outcomes and achievements:

```
PS C:\Users\shard\.vscode\Projects\JUIT-GPT> python JuitGPT.py
Found model file at  models/ggml-gpt4all-j-v1.3-groovy.bin
gptj_model_load: loading model from 'models/ggml-gpt4all-j-v1.3-groovy.bin' - please wait ...
gptj_model_load: n_vocab = 50400
gptj_model_load: n_ctx   = 2048
gptj_model_load: n_embd  = 4096
gptj_model_load: n_head  = 16
gptj_model_load: n_layer = 28
gptj_model_load: n_rot   = 64
gptj_model_load: f16     = 2
gptj_model_load: ggml ctx size = 5401.45 MB
gptj_model_load: kv self size  =  896.00 MB
gptj_model_load: ..................................... done
gptj_model_load: model size =  3609.38 MB / num tensors = 285
```

Fig: 5.1 Loading Models

```
Enter a query: Tell me about the campus of juit
 The JUIT is a university located in Himachal Pradesh, India. It was established by Act No 14 of
2002 and provides undergraduate degree programs in various fields such as Bioinformatics, Biotech
nology, Civil Engineering, Computer Science & Engineering, Electrical Engineering, Electronics &
Communication, Information Technology, Mechanical Engineering, Physics, Statistics, Zoology, Bota
ny, Chemistry, Geography, Economics, Political Science, Sociology, Psychology, History and Englis
h. The campus is spread over an area of approximately 1,000 acres with a population of around 10,
000 students. It has modern facilities such as computer labs, libraries, laboratories, sports gro
unds, cafeteria, auditoriums, hostels for boys & girls, medical center, post office, bank branche
s, etc.
```

Fig: 5.2 Entering query and getting answer

5.1.1 DIVERSITY IN USER QUERIES:

The diverse nature of user queries, ranging from academic inquiries to administrative details and event-related information, reflects the model's versatility. JUIT-GPT has successfully navigated and responded to queries across various domains, showcasing its ability to cater to the multifaceted information needs of users.

5.1.2 EFFECTIVE USER QUERY HANDLING:

JUIT-GPT successfully handles a variety of user queries, including ambiguous, complex, and context-dependent questions. The model's natural language processing capabilities and adaptive algorithms contribute to improved query resolution and a more intuitive user interaction.

5.1.3 EDUCATIONAL RESOURCE OPTIMIZATION:

The model has become an educational empowerment tool, assisting students in accessing course information, faculty details, and relevant academic resources with unprecedented ease. Faculty members also benefit from streamlined access to administrative information, contributing to more efficient academic management.

5.1.4 ADMINISTRATIVE EFFICIENCY:

JUIT-GPT has streamlined administrative processes by providing quick and accurate responses to administrative announcements, event details, and other relevant information queries. This efficiency improvement translates into time and resource savings for administrative staff, allowing them to focus on more strategic aspects of their roles.

5.1.5 RESEARCH AND DEVELOPMENT OPPORTUNITIES:

The success of JUIT-GPT has opened up avenues for further research and development within the field of natural language processing (NLP) and AI-driven information systems. Collaborations with academic and industry partners are being explored to contribute to the broader landscape of language models and intelligent information retrieval.

CHAPTER 6

CONCLUSION AND FUTURE SCOPE

6.1 CONCLUSION

6.1.1 RECAP OF OBJECTIVES AND CONTRIBUTION

In this full-scale journey, the main goals from the start were carefully inspected and broken down. This research dove into different areas of natural language processing, including sentiment analysis, machine translation, name entity recognition, among others. This deep research shared many discoveries and added value to the bigger picture of language models and sentiment interpretation. It also shed light on the issues within the world of NLP.

6.1. 2 SUMMARY OF MAJOR FINDINGS AND TREND

A mix of various discoveries from different areas has shown trends and patterns in natural language processing. We've seen where sentiment analysis models need improvement and identified the importance of strong speech recognition in many types of sound settings. Stories are coming together that show the challenges in natural language processing (NLP) but also its potential routes for progress.

6.1.3 THE SIGNIFICANCE OF RESEARCH OUTCOME

This study's results carry strong messages for schools and businesses. They give clues for more research, bringing light to areas not yet studied enough in the big world of using computers to understand language. Plus, these facts could change things like figuring out feelings from words, understanding spoken words, translating languages, and more in real-world uses.

6.1.4 SYNTHESIS OF RESEARCH THEME AND PATTERN

Combining different topics and trends seen throughout the studies highlights how everything is linked in the world of natural language processing. Bringing together these results shows a pattern filled with new ideas, hurdles, and chances. It stresses the importance of using well-rounded strategies to tackle the difficulties present in language-focused tasks.

6.1.5 ALIGNMENT WITH RESEARCH OBJECTIVE

This study provides beneficial data for the academic world, bringing fresh views and understanding. It aims to advance the study of NLP by revealing less-studied aspects. This work strives to ignite academic debates, encouraging more investigation, feedback, and improvement within the discipline.

6.1.6 CONTRIBUTION TO THE FIElD AND ACADEMIC DISCOURSE

Rephrase discovering new areas usually raises fresh inquiries. This research opens the door for upcoming scholars to probe more into delicate parts of feelings analysis, name identification, machine interpretation, and more. The uncovered gaps act as guides for future attempts, urging scholars to discover, enrich, and resolve current constraints.

6.1.7 FUTURE DIRECTION FOR RESEARCH

This report turns study results into easy-to-follow tips. It gives advice for real life use. It talks about improving ways to understand feelings from data and building stronger voice recognition systems. This research helps more than just schools. It changes the field of technology too.

6.1.8 RECOMMENDATION FOR PRACTICAL APPLICATION

The methods used in this study act like maps, helping us understand the tricky landscape of natural language processing. We mixed number-based studies, detailed evaluations, and contrast analyses to get a complete view of the details in sentiment analysis, machine

translations, and others. Thinking about the methods shows what and where they are strong and weak, and opens up ways to improve or take new paths in future research.

6.1.9 REFLECTING ON METHODOLOGY

Understanding the challenges faced in our study is vital to promoting an atmosphere of ongoing improvement and expansion. While our science was strong and wide-ranging, issues with getting enough data, model flexibility, or shifting contexts remain. In these hitches, however, limitless chances for discovering more are hidden. These issues invite more and more in-depth studies, encourage innovation, and challenge the vast field of natural language processing to increase our knowledge.

6.1.10 ACKNOWLEDGEMENT LIMITATION AND OPPORTUNITY

The mixed-up world of natural language processing means we need teamwork from many areas as. Making linguists, computer geeks, brain experts, and other professionals work together can really push NLP research further. By stirring up group work, this study wants to create powerful conversations and mash-ups. Using ideas from lots of areas to start new plans, ways of working, and big steps in figuring out, studying, and making sense of human language.

6.1.11 CONCLUDING REMARKS AND REFLECTION

To wrap things up, our careful study of natural language processing has revealed complex aspects, obstacles, and potential. The path we took shed light on undiscovered areas, making way for more look into, creativity, and use. As the carriers of wisdom, this study encourages other researchers, professionals, and fans to walk the trails we found, aiming for top-notch work, innovation, and progress in the continually shifting world of natural language processing.

6.2 FUTURE SCOPE

The destiny scope of this project marks an evolution towards leveraging the Dart framework to develop robust packages and web solutions. Additionally, a concerted attempt may be directed in the direction of augmenting the dataset used for education and trying out functions, aiming for a more numerous and complete corpus. Alongside dataset enlargement, the paramount goal remains the enhancement of model accuracy throughout numerous domain names inside natural language processing.

The adoption of the Dart framework for software and web improvement stands as a pivotal step in the direction of creating seamless, efficient, and person-centric interfaces. Leveraging Dart's versatility, this future undertaking objectives to craft applications with enhanced person reviews, leveraging the framework's competencies for go-platform compatibility, fast development, and scalability.

Efforts closer to dataset growth might be paramount, aiming to encompass a more extensive and diverse variety of textual statistics from diverse assets, domains, and languages. This expansion initiative seeks to seize a broader spectrum of linguistic nuances, contexts, and expressions, fostering a more complete information inside herbal language processing tasks.

Diversifying the dataset involves the inclusion of multilingual text, domain-unique corpora, and sundry textual formats. Incorporating statistics from various assets such as social media, clinical literature, conversational communication, and greater will enhance the dataset's richness, permitting fashions to adapt to numerous linguistic landscapes.

Alongside dataset enlargement, dedicated efforts may be directed in the direction of meticulous facts annotation and best guarantee measures. Rigorous annotation protocols and quality assessments will ensure the dataset's reliability, accuracy, and alignment with particular assignment objectives, laying the basis for sturdy model education and evaluation.

With an accelerated and subtle dataset, the focus shifts in the direction of improving version architectures, optimizing hyperparameters, and exploring superior education strategies. Iterative version refinements will aim to raise accuracy, robustness, and generalization abilties across diverse herbal language processing duties.

Augmenting contextual embeddings and representations within models will be a key region of exploration. Strategies regarding transformer architectures, interest mechanisms, and contextual expertise might be hired to decorate models' abilities to capture intricate linguistic contexts and nuances.

Tailoring models to domain-particular responsibilities through pleasant-tuning approaches will be prioritized. Fine-tuning techniques, which includes switch learning paradigms and area edition techniques, will empower models to excel in specialized domains, catering to nuanced language information within precise industries or applications.

A concerted attempt will be made to deal with moral considerations and mitigate biases inside the dataset and models. Ethical AI practices, equity tests, and bias detection mechanisms may be incorporated to ensure equitable and accountable deployment of NLP fashions.

The established order of comprehensive evaluation metrics and benchmarking requirements remains important. Robust evaluation methodologies can be devised to gauge version performance, ensuring consistency and comparability throughout experiments and benchmarks.

REFERENCES

[1] S.Y.Chien, G.J.Hwang,“A research proposal for an AI chatbot as virtual patient agent to improve nursing students’ clinical inquiry skills” , 2023,ICAIE 2023.

[2] X.Chen, G.Cheng, D.Zou, B.Zhong, H.Xie. “Artificial Intelligent Robots for Precision Education”,2023,Educational Technology Solution.

[3] T.C.Hsu, H.L.Huang , G.J.Hwang, M.S.Chen, “Effects of Incorporating an Expert Decision-making Mechanism into Chatbots on Students’ Achievement, Enjoyment, and Anxiety”.2023, Educ. Technol. Soc.

[4] A.Tlili, B.Shehata, M.A.Adarkwah, A.Bozkurt, D.T.Hickey, R. Huang, B.Agyemang, ”What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education”. Smart Learn. Environ. 2023.

[5] M.Murtaza, Y.Ahmed, J.A.Shamsi, F.Sherwani,M.Usman,“AI-Based Personalized E-Learning Systems”,2022, Challenges, and Solutions. IEEE Access.

[6] O.Tapalova, N.Zhiyenbayeva “Artificial Intelligence in Education: AIEd for Personalized Learning Pathways”,2022 Electron. J. e-Learn.

[7] M.A.Kuhail, N.Alturki, S.Alramlawi, K.Alhejori,“Interacting with educational chatbots: A systematic review.”,2022, Educational Information Technology .

[8] K.Zhang, A.B.Aslan, “AI technologies for education: Recent research and future directions”,2021, Computer Education. Artificial Intelligence.

[9] V.González-Calatayud, V.Prendes-Espinosa, P. Roig-Vila, “Artificial Intelligence for Student Assessment: A Systematic Review”.2021,Appl. Sci.

[10] A.Følstad, C.Taylor, “Investigating the user experience of customer service chatbot interaction: A framework for qualitative analysis of chatbot dialogues.”,2021, Qual. User Exp.

[11] Y.Windiatmoko, R.Rahmadi and A.F.Hidayatullah, “Developing Facebook Chatbot Based on Deep Learning Using RASA Framework for UniversityEnquiries”,2021,Materials Science And Engineering.

[12] L.Chen, P.Chen, Z.Lin, “Artificial Intelligence in Education: A Review”,2020, IEEE Access.

[13] X.Chen, Xie, H.,Zou, D.; Hwang, G.-J. Application and theory gaps during the rise of Artificial Intelligence in Education. Computer Education Artificial Intelligence 2020.

[14] T.Araujo, “Conversational agent research toolkit: An alternative for creating and managing chatbots for experimental research”,2020,Computer Communication. Res.

[15] S.N.M.S.Pi and M.A.Majid.“Components of Smart Chatbot Academic Model for a University Website”,2020, Emerging Technology in Computing, Communication, and Electronics (ETCCE) IEEE.

[16] O.Gbenga, T.Okedigba and H.Oluwatobi, “An Improved Rapid Response Model for University Admission Enquiry System Using Chatbot”,2020 Int. J. Comput.

[17] S.Kumari, Z.Naikwadi, A.Akole and P.Darshankar, “Enhancing College Chatbot Assistant with the Help of Richer Human-Computer Interaction and SpeechRecognition”. 2020,International Conference On Electronics and Sustainable Communication Systems (ICESC). IEEE.

[18] H.Koundinya, A.K.Palakurthi, V.Putnala and A.Kumar, “Smart College Chatbot using ML and Python”,2020 International Conference on System, Computation, Automation, and Networking (ICSCAN), IEEE.

[19] N.P.Patel, D.R.Parikh, D.A.Patel and R.R.Patel “AI and web-based human-like interactive university chatbot(UNIBOT)”,2019, InternationalConference on Electronics, Communication and Aerospace Technology (ICECA)IEEE.

[20] Y.K.Dwivedi, L.Hughes, E.Ismagilova, G.Aarts, C.Coombs, T.Crick, Y.Duan, R.Dwivedi, J.Edwards, A.Eirug, “Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research,”,2019, practice and policy. Int.J.Inf. Manag.

[21] P.Kulkarni, A.Mahabaleshwarkar, M.Kulkarni, N.Sirsikar, and M.Gadgil, “Conversational AI: An Overview of Methodologies, Applications & Future Scope”,2019,International Conference On Computing, Communication, Control And

Automation (ICCUBEA). IEEE.

[22] K.Ralston, Y.Chen, H.Isah and F.Zulkernine. “A voice interactive multilingual student support system using IBM Watson”.2019, IEEEInternational Conference On Machine Learning And Applications (ICMLA) .

[23] M.Phillips, H.Marsden, W.Jaffe, R.N.Matin, G.N.Wali, J.Greenhalgh, E.McGrath, R.James, “Assessment of Accuracy of an Artificial Intelligence Algorithm to Detect Melanoma in Images of Skin Lesions”.2019, JAMA Network.

[24] W.Xiong, L.Wu, F.Alleva, J.Droppo, X.Huang, and A.Stolcke, “The Microsoft 2017 conversational speech recognition system”.2018,IEEE international conference on acoustics, speech and signal processing (ICASSP).IEEE.

[25] D.Braun, A.H.Mendez, F.Matthes, and M.Langen, August. “Evaluating natural language understanding services for conversational question answering systems”. 2017,In Proceedings of the 18th AnnualSIGdial Meeting on Discourse and Dialogue.

PLAGIARISM

ORIGINALITY REPORT

2%	1%	1%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Akshay Kulkarni, Adarsha Shivananda, Anoosh Kulkarni, Dilip Gudivada. "Applied Generative AI for Beginners", Springer Science and Business Media LLC, 2023 Publication	1%
2	blog.pangeanic.com Internet Source	<1%
3	www.healthcare.digital Internet Source	<1%
4	chataigpt.org Internet Source	<1%
5	Koushik Chandra Sarker, Md. Mijanur Rahman, Abu Siam. "Anglo-Bangla Language-Based AI Chatbot for Bangladeshi University Admission System", 2023 International Conference on Communications, Computing and Artificial Intelligence (CCCAI), 2023 Publication	<1%
6	de.readkong.com Internet Source	<1%

7	healthdocbox.com Internet Source	<1%
8	eprints.utem.edu.my Internet Source	<1%
9	www.mdpi.com Internet Source	<1%
10	www.researchgate.net Internet Source	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

PLAGIARISM VERIFICATION REPORT

Date: …….

Type of Document (Tick): PhD Thesis | M.Tech Dissertation/ Report | B.Tech Project Report | Paper

Name: _____ __**Department:** _____ **Enrolment No** _____

Contact No. _____**E-mail.** _____

Name of the Supervisor: _____

Title of the Thesis/Dissertation/Project Report/Paper (In Capital letters): _____

_____

_____

UNDERTAKING

I undertake that I am aware of the plagiarism related norms/ regulations, if I found guilty of any plagiarism and copyright violations in the above thesis/report even after award of degree, the University reserves the rights to withdraw/revoke my degree/report. Kindly allow me to avail Plagiarism verification report for the document mentioned above.

Complete Thesis/Report Pages Detail:

- Total No. of Pages =
- Total No. of Preliminary pages =
- Total No. of pages accommodate bibliography/references =

(Signature of Student)

FOR DEPARTMENT USE

We have checked the thesis/report as per norms and found **Similarity Index** at ……..(%). Therefore, we are forwarding the complete thesis/report for final plagiarism check. The plagiarism verification report may be handed over to the candidate.

(Signature of Guide/Supervisor) **Signature of HOD**

FOR LRC USE

The above document was scanned for plagiarism check. The outcome of the same is reported below:

<table>
<tr><th>Copy Received on</th><th>Excluded</th><th>Similarity Index (%)</th><th colspan="2">Generated Plagiarism Report Details (Title, Abstract & Chapters)</th></tr>
<tr><td rowspan="2"></td><td rowspan="4">• All Preliminary Pages
• Bibliography/Images/Quotes
• 14 Words String</td><td rowspan="2"></td><td>Word Counts</td><td></td></tr>
<tr><td rowspan="2">Report Generated on</td><td>Character Counts</td><td></td></tr>
<tr><td rowspan="2"></td><td>Submission ID</td><td>Total Pages Scanned</td><td></td></tr>
<tr><td></td><td>File Size</td><td></td></tr>
</table>

Checked by
Name & Signature **Librarian**

……

Please send your complete thesis/report in (PDF) with Title Page, Abstract and Chapters in (Word File) through the supervisor at plagcheck.juit@gmail.com