International Journal of Scientific Research and Engineering Development-– Volume 9 Issue 2, Mar-Apr 2026
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Smart Campus Conversational Assistant An Intelligent System for College Queries

Zoya Anjum[1],Smanavi Reddy[2],Geetha Sri[3],Bazila Wahaj[4]
[1](Department of Computer Science and Engineering, Stanley College of Engineering, Hyderabad, India zoyaanjum203@gmail.com)
[2] (Department of Computer Science and Engineering, Stanley College of Engineering, Hyderabad, India Vakitismanavi@gmail.com)
[3](Department of Computer Science and Engineering, Stanley College of Engineering, Hyderabad, India geethasrimamilla19189@gmail.com)
[4](Department of Computer Science and Engineering, Stanley College of Engineering, Hyderabad, India bazila@stanley.edu.in)

----------**********----------

Abstract

Colleges and other higher education institutions are increasingly dependent on online resources for academic and administrative data. Although these technologies combine institutional resources, they frequently employ keyword-based search methods and predetermined content formats, which restricts the user's ability to access information rapidly. As a result, students occasionally put off looking for the information they need. These architectural constraints also limit typical rule-based chatbots that are used on these platforms, and that rely on predetermined patterns and a lack of in-depth semantic understanding. As a result, their academic achievement is negatively impacted by their inability to comprehend questions that are context-dependent or use dynamic language. In order to address these issues, this study suggests integrating a Smart Campus Conversational Assistant onto a university website to enable intelligent, domain-specific access. Semantic vector-based retrieval and regulated generative language modelling are combined in the system's Retrieval-Augmented Generation (RAG) design. Before being indexed by Facebook FAISS-powered semantic vector search for speedy vector similarity matching, the Sentence Transformer model preprocesses and encodes institutional knowledge into dense vector representations. In order to find relevant information, incoming queries are buried inside the same vector space throughout runtime. The results are then sent to a Large Language Model (LLM) for logical and well-documented responses.
The suggested hybrid retrieval-generation method increases the level of semantic relevance as well as context alignment through a reduction in false positives, thereby enhancing the transmission of reliable information through intelligent campus networks.

Keywords — Smart Campus,Retrieval-Augmented Generation (RAG), Semantic Search,Information Retrieval,FAISS, Conversational Assistant,Large Language Models.

----------**********----------

I. INTRODUCTION

Web-based gateways are becoming increasingly important for universities to manage and disseminate administrative and academic data.Even though these websites collect institutional data, their keyword-based search engines and hierarchical navigation tools limit the ability to locate relevant information [6]. As a result, consumers are frequently required to input certain keywords or browse through numerous pages to locate the information they are looking for, which causes delays and lowers user satisfaction [18]. To increase accessibility, many colleges have incorporated chatbot-based interfaces into their portals [7], [8]. However, traditional rule-based chatbots operate on predetermined patterns and scripted responses, which restricts their ability to answer questions that use dynamic language or semantic diversity [7]. Statistical similarity approaches such as Term Frequency Inverse Document Frequency (TF-IDF) offer incremental improvements in text matching [6], yet they remain dependent on surface-level word overlap rather than true semantic understanding [17]. These limitations are particularly critical in institutional environments where information related to admissions, fee structures, eligibility criteria, and academic Laws must remain authoritative, dependable, and trustworthy. Recent advances in transformer-based Large Language Models (LLMs) have demonstrated significant advancements in the production and comprehension of natural language [1], [2]. Standalone models are prone to hallucinations, as they are not backed by any certified source of information, which might result in contextually aware as well as fluent but factually weak responses [3]. Unrestricted generative systems are not suitable

in critical academic knowledge sites due to the aforementioned risks. In order to deal with the aforementioned challenges, this study suggests a Smart Campus Conversational Assistant based on a structured, multi-layered processing system [14].

The Retrieval-Augmented Generation (RAG) model, upon which this study is based, combines regulated generative modeling with semantic retrieval based on vector models [3]. Knowledge stored in a structured format at the institution level is converted into rich semantic vectors through a pre-trained Sentence Transformer model [4]. These vectors are then indexed using the Facebook AI Similarity Search (FAISS) algorithm, which allows for fast similarity-based retrieval [5]. User queries are then converted into the same space to retrieve context based information during inference, and then input into a Large Language Model to produce logical, context-aware, and accurate information. This approach not only increases the level of semantic relevance but also helps in the reduction of hallucinations. Furthermore, the combination of deterministic, statistical, and semantic retrieval methods into a single hierarchical structure increases the reliability of queries as well as the accuracy of domain-specific information compared to conventional methods of chatbots [7], [8]. For intelligent information access in modern smart campus ecosystems, this approach provides a scalable, dependable, and efficient solution.

II. A LITERATURE REVIEW

Intelligent chatbot systems for the academic and institutional environment can be developed with the recent advancements in domain of interactive AI communication systems and automated text processing techniques.Studies on the application of contemporary NLP technologies were first presented with the Transformer approach, which has enhanced the contextual link recognition capacity of the proposed models [1]. This has allowed the application of BERT to generate bidirectional language representations, which has Optimized the functionality of the several Computational language processing applications and laid the foundation for the development of several Intelligent virtual assistants [2]To improve the functionality of language applications that depend on knowledge to a significant extent, the Knowledge-augmented generation hybrid approach has been proposed, which utilizes External knowledge-assisted language models. By doing so, it reduces hallucinatory results and thus facilitates systems in providing context-dependent accurate results [3]. Conversational AI systems have also advanced significantly due to the use of embedding in semantic representation techniques. By providing efficient sentence-level embeddings that enable semantic similarity comparisons between queries and entries in the stored knowledge base, Sentence-BERT streamlines information retrieval in chatbot applications [4]. Efficient similarity search techniques like FAISS have been widely used to enable large-scale vector search operations, allowing for the rapid retrieval of semantically connected information from huge knowledge bases [5].

Statistical methods like Frequency–inverse frequency weighting have employed extensively in conventional information retrieval systems for text data representation and similarity computation between queries and documents. All these approaches offer a simple but useful framework for document retrieval and query matching systems [6].Several researchers have also examined the use of chatbots in school systems. An automatic conversation system may be useful for helping students access information on admissions, courses, and university services, as in the case of a chatbot for assisting students with college websites [12].However, the system itself did not have advanced semantic understanding abilities [7], and most of its answers were pre-defined. Further research led to the design of intelligent campus consultation systems that use natural language processing to automatically handle institutional queries and improve user experience.The systems continued to face problems of contextual consistency and were trying to process dynamic language [8], despite their high usability. It has also been suggested that the services provided by university websites may be supplemented by AI-driven chatbot frameworks using machine learning algorithms [15], [16].These systems would still be depending on certain prior set training data and knowledge bases [9] despite the improvement in the response automation and user interface. The use of the RAG-based conversational systems has been researched in recent years for use in the educational support programs such as test preparation systems. This has shown the potential for improving response accuracy through the integration related to retrieval and generative models while maintaining the relevancy of the responses [10]. To aid in the acquisition of administrative and student information, assistants based on the RAG model have been developed. This has provided the utilization of semantic embeddings and data bases within improving accuracy for the responses in the academic field. Further providing the need for constant campus assistance [16], automated question processing through COLLEGIA-like conversational systems has been developed. This has highlighted the need to leverage the technology in the field of higher education and the concerns surrounding the breadth and reliability of the knowledge base for the chatbots [12], [16].

Despite the hope offered in these studies for the applications of conversational AI in educational settings, the majority of the current solutions are either based on rules or lack a controlled integration of methods for retrieval and generation models [15]. To overcome these limitations, the present study advises a hierarchical architecture incorporating statistical similarity analysis, deterministic intent matching, and embedding-based semantic retrieval to provide exact and domain-specific academic information access for intelligent campus systems. The table below shows conventional and RAG models given in table1.

Feature / Metric	Traditional Chatbot	RAG-Based Chatbot

Knowledge Source	Fixed FAQ / Script	Dynamic retrieval from documents / database
Handling Paraphrased Queries	Poor	High
Hallucination	High	Low
Out-of-Domain Filtering	Poor	High

Table 1: Comparison of Traditional vs RAG-based Chatbots

III. SYSTEM ARCHITECTURE

A. Overall Hybrid Architecture

Statistical similarity analysis, semantic vector search, deterministic processes, and the generation of controlled responses will be used in the future form of the Smart Campus Conversational Assistant as part of a hierarchical domain-specific retrieval-augmented generation approach to enable users to write responses. This approach is based on precision and reliability through careful consideration of user requests at different decision levels before creative thinking is applied. Fig1 shows the design of system.

Web Chatbot Interface is part of the Presentation Layer of the institutional portal's architecture. The application layer has hierarchical query processing modules for statistical similarity analysis, deterministic intent matching, and semantic retrieval. Part of the Data Layer is the well-developed knowledge base of the institution and the vector index based on the library FAISS. Only when particular semantic confidence criteria are satisfied is the Model Layer, which contains the local Large Language Model (LLM), triggered. This multilayer approach ensures efficiency by restricting the kinds of queries that can be handled.

B. Hierarchical Process Flow for Query

To process dynamic requests, a confidence-based processing procedure is followed, which will implement a series of hierarchy levels such as the evaluation of trustworthiness through deterministic, statistical, and semantic evaluation methods.The evaluation of a query will first require the input to be normalised and validated prior to using the deterministic intent matching module to ascertain if the question can be matched with any existing pre-defined queries for an institution.If a match exists, then the answer will be returned immediately. If no match exists, the degree of statistical significance will be determined using the application of CS.All requests whose statistics do not meet sufficient confidence intervals are then routed to the Semantic Retrieval Layer.

The Semantic Retrieval Layer uses FAISS-based similarity searches and creates dense embedded representations of each data component to find contextually relevant pieces of data. If a request does not meet the specified amount of similarity within context, the system will reject the request. If it does meet that level of similarity, the system will then invoke Retrieval-Augmented Generation. Figure 2 shows an example of how these processes will flow.

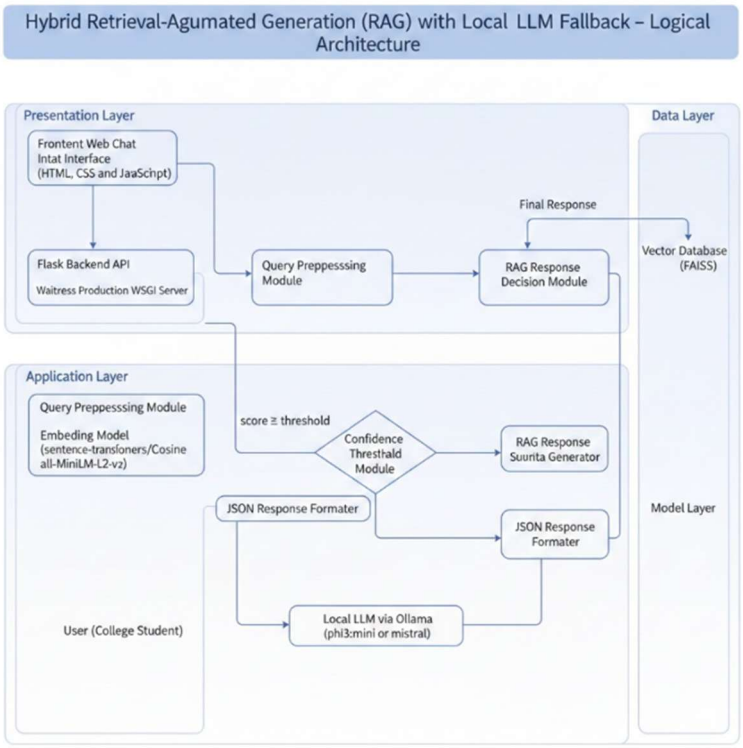

Fig. 1. Overall Hybrid RAG-Based System Architecture.

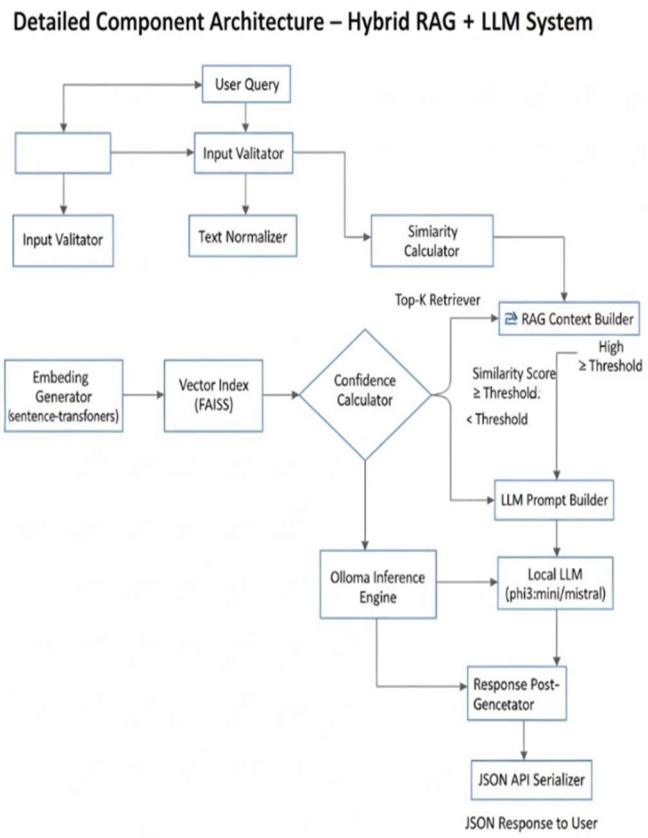

Fig. 2. Detailed Component-Level Processing with Confidence-Based RAG Invocation.

C. Deployment Overview

Local operation uses a Flask-based backend, a FAISS vector store, and a local LLM environment. This deployment configuration ensures cost efficiency and data privacy by staying independent of external API services and providing real-time conversational performance.

IV. PROPOSED METHODOLOGY

The suggested system uses a hybrid, confidence-driven retrieval technique that combines semantic embedding-based retrieval, statistical similarity analysis, and deterministic intent matching within a regulated RAG framework. The goal is to avoid hallucinatory answers while maintaining domain-specific correctness.

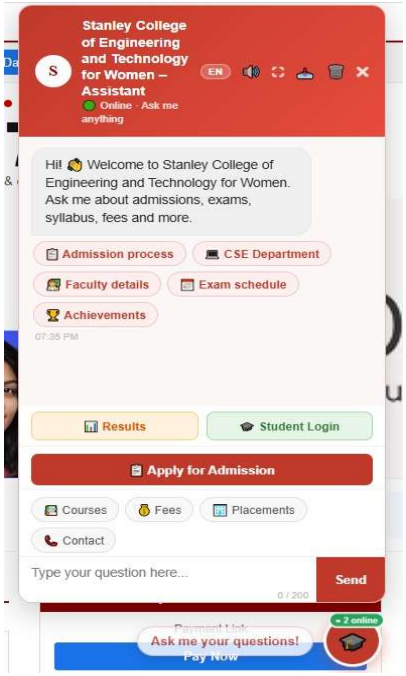

Fig. 3. Web-based Smart Campus Conversational Assistant of institutional portal.

A. *Knowledge Base Construction*

Ordered JSON structure maintains institutional data consisting of confirmed question-answer pairs. Each query is converted during the offline phase into a dense embedding vector.These vectors are indexed using FAISS to enable efficient similarity search

B. *Hierarchical Query Processing*

1) Deterministic Layer: In the first stage, rule-based intent matching is applied to handle structured institutional queries such as admissions, faculty information, and examination results. This layer uses predefined patterns and keyword matching to quickly identify known query types and return deterministic responses.

$$Qi \rightarrow vi \in Rd$$

2) Statistical Layer: If the query is not resolved in the deterministic layer, a statistical similarity approach is applied. The query and stored documents are transformed using TF–IDF vectorization, and the similarity between vectors is computed using cosine similarity to retrieve the relevant information

$$sim(v\_i, d\_j) = (v\_i \cdot d\_j) / (||v\_i|| \, ||d\_j||)$$

3) Semantic Level: Finally, the sentence embedding approach transforms inquiries into thick vector embeddings. These embeddings are matched to the indexed document vectors saved in the FAISS database to locate semantically related data. The resemblance between queries embedding and stored embeddings is assessed if the similarity score reaches the threshold requirements.

$$Sinsemantic \geq \theta 2$$

C. Controlled Retrieval-Augmented Generation

Based only on institutional information, the retrieved contextual data is supplied to locally hosted Large Language Model to generate replies. By including the gathered information as contextual input, the system preserves the relevance and correctness of the responses to the field generated. Queries not meeting the criteria specified are quickly rejected to avoid out-of-domain responses and reduce the likelihood of hallucinatory findings.

D. Dataset Description

The data for this study is institution-specific academic data taken from official college sources, such as admission procedures, qualifying criteria, fee schedules, exam regulations, and frequently asked questions. The gathered data was stored in JSON format after being arranged into organized question-answer pairs. To enable semantic similarity searches, every query item in the dataset was changed into a vector representation in the offline pre-processing stage. Using the ordered dataset as its knowledge base, the suggested conversational assistant service can retrieve relevant information and provide relevant answers to user questions.

V. EXPERIMENTAL RESULTS AND ANALYSIS

The proposed hybrid Retrieval-Augmented Generation based Smart Campus Conversational Assistant was evaluated in terms of its domain-restricted behaviour, semantic strength, and response accuracy. The hierarchical query processing method as well as the controlled generative response mechanism therefore shown.

A. Evaluation Setup

Representative institutional problems pertaining to admissions, tuition rates, eligibility requirements, academic standards, and testing techniques were used to assess the system. Resilience was evaluated using the dataset by paraphrasing inputs, posing questions from outside the field, and sequentially querying.

B. Performance Evaluation

With minimal processing burden, the deterministic layer rapidly handled sorted requests. Statistical similarity made possible by vector analysis based on TF-IDF improved lexical matching. Moreover, the embedding-based semantic retrieval layer helped for contextual understanding of queries with many variations in natural language and phrasing. The hierarchical approach of processing gave efficiency top priority by only engaging more complex retrieval levels when needed.

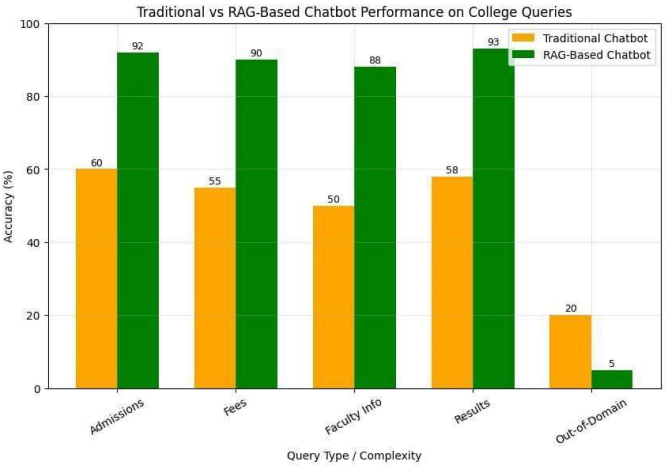

Fig.4. Comparison of Performance between traditional chatbot and RAG-based chatbot.

C. Domain Constraint Enforcement

Queries that do not fulfil the predetermined standards for semantic similarity are not, however, entered into the generative model. A well-regulated fallback response directs users to the supported institutional subjects. This preserves the domain integrity and lowers the likelihood of skewed or irrelevant outcomes.

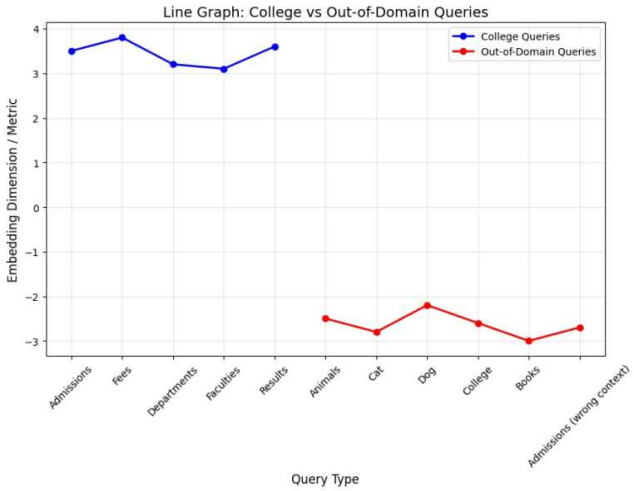

Fig. 5. Separation of out-of-domain queries and institutional queries

D. System Efficiency

The FAISS-based indexing allowed for a rapid runtime semantic search in the document embedding space. To reduce the computational burden while guaranteeing the precision and pertinence of the answer, the tiered architecture design made sure that the Large Language Model (LLM) was used only when necessary.

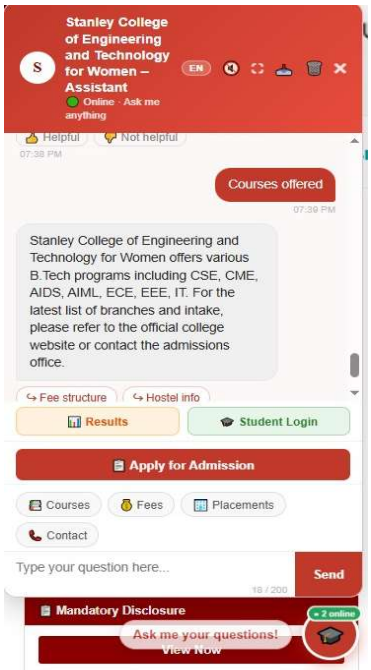

Fig. 6. Example query showing the chatbot providing information about courses offered by the institution.

VI. CONCLUSION

In this study, we introduced a Smart Campus Conversational Assistant for retrieving academic information for a particular domain by applying a hybrid Retrieval-Augmented Generation (RAG) system. The suggested system combines statistical similarity evaluation, embedding-based semantic retrieval, and deterministic intent matching within a confidence-driven hierarchical architecture to guarantee precise and trustworthy query management. The framework efficiently reduces hallucinated outputs while preserving contextual consistency and semantic relevance by anchoring generative replies in institution-specific knowledge indexed via FAISS. Additionally, the domain-restricted processing system protects institutional data integrity by guaranteeing the safe handling of queries that are unsupported or outside of the scope.

Compared to standard rule-based and keyword-based chatbot systems, experimental evaluation revealed enhanced query relevance, regulated generative behaviour, and effective response performance. The layered processing approach reduces unnecessary calculation as well as enables flexible deployment in institutional contexts.

Suggested framework provides a realistic and cost-effective way to improve information access in intelligent campus environments. To further increase flexibility and system responsiveness, future improvements may concentrate on better integration with real-time institutional databases and regulated growth of knowledge repositories.

ACKNOWLEDGMENT

The authors express their gratitude to our faculty mentors and our project guide, whose valuable guidance, suggestions, and feedback helped make the successful completion of this research possible. Secondly, the authors like to extend our sincere gratitude to our college and departmental staff, whose support in providing the necessary resources, infrastructure, etc. were greatly instrumental in the successful execution of the research project Lastly, we acknowledge the valuable contributions of our team members, whose teamwork was greatly instrumental in the successful completion of this research.

REFERENCES

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Proceedings of NAACL-HLT*, 2019. [Online]. Available: https://doi.org/10.48550/arXiv.1810.04805

[2] A. Vaswani *et al.*, "Attention Is All You Need," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017. [Online]. Available: https://doi.org/10.48550/arXiv.1706.03762

[3] J. Johnson, M. Douze, and H. Jégou, "Billion-Scale Similarity Search with GPUs," *IEEE Transactions on Big Data*, vol. 7, no. 3, pp. 535–547, 2021. [Online]. Available: https://doi.org/10.1109/TBDATA.2019.2921572

[4] G. Salton and C. Buckley, "Term-Weighting Approaches in Automatic Text Retrieval," *Information Processing & Management*, vol. 24, no. 5, pp. 513–523, 1988. [Online]. Available: https://doi.org/10.1016/0306-4573(88)90021-0

[5] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," in *Proc. EMNLP-IJCNLP*, 2019. [Online]. Available: https://doi.org/10.48550/arXiv.1908.10084

[6] P. Lewis *et al.*, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. [Online]. Available: https://doi.org/10.48550/arXiv.2005.11401

[7] A. Adamopoulou and L. Moussiades, "An Overview of Chatbot Technology," in *Artificial Intelligence Applications and Innovations*, Springer, 2020.

[8] J. Hill, W. R. Ford, and I. G. Farreras, "Real Conversations with Artificial Intelligence: A Comparison Between Human–Human and Human–Chatbot Conversations," *Computers in Human Behavior*, vol. 49, pp. 245–250, 2015.

[9] S. Shum, X. He, and D. Li, "From ELIZA to XiaoIce: Challenges and Opportunities with Social Chatbots," *Frontiers of Information Technology & Electronic Engineering*, vol. 19, no. 1, pp. 10–26, 2018.

[10] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*. Sebastopol, CA, USA: O'Reilly Media, 2009.

[11] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, "SQuAD: 100,000+ Questions for Machine Comprehension of Text," in *Proc. EMNLP*, 2016.

[12] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," arXiv:1301.3781, 2013.

[13] J. Weizenbaum, "ELIZA—A Computer Program for the Study of Natural Language Communication Between Man and Machine," *Communications of the ACM*, vol. 9, no. 1, pp. 36–45, 1966.

[14] D. Jurafsky and J. H. Martin, *Speech and Language Processing*, 3rd ed. Stanford University, 2023.

[15] D. Patel, N. Shetty, P. Kapasi, and I. Kangriwala, "Design and Research of Intelligent Chatbot for Campus Information Consultation Assistant," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, 2023.

[16] R. Thamilselvan *et al.*, "Developing an AI-Driven Chatbot for Enhanced College Website Support Using Machine Learning," in *Proc. International Conference on Expert Clouds and Applications (ICOECA)*, Bengaluru, India, 2024, pp. 719–726.

[17] N. P. *et al.*, "RAG-Based AI Chatbot for Student and Institutional Assistance," *IJRASET*, 2025.

[18] L. Gkinko and A. Elbanna, "AI Chatbots Sociotechnical Research: An Overview and Future Directions," *IEEE Access*, 2025..

