Chatbot-Enhanced Placement Data Management and Summarization Application

Prof. Y.A. Handge[1], Tejal Kolhe[2], Maithili Mandla[3], Dnyaneshwary Dolas[4], Shreya Mali[5]

[1-2-3-4-5]*Computer Engineering SCTR's Pune Institute of Computer Technology* Pune, India

Abstract—**This project focuses on improving the management and accessibility of placement data for students through the use of automated data collection and advanced natural language processing (NLP) techniques. Data will be collected from seniors who have secured placements via a simple form that requires minimal information, such as their email and the name of the company. A text summarizer will then condense this information to highlight key insights, making it easier for users to understand. Additionally, a chatbot, trained on the gathered data, will provide real-time support by responding to inquiries about various com- panies. Both the chatbot and the text summarizer are designed to effectively address this specific use case.**

I. INTRODUCTION

In today's digital landscape, the rapid growth of textual data poses significant challenges for effective processing and utilization [1]. This challenge is especially relevant in educational environments, where students preparing for job placements must sift through extensive information regarding company recruitment processes, experiences shared by seniors, and interview strategies. Historically, managing this data has relied on manual methods, which can be both inefficient and time-consuming. The increasing demand for more efficient approaches to process such information has prompted the in- vestigation of automated systems that can enhance the collec- tion and accessibility of pertinent insights [2]. Consequently, this project seeks to refine the management and access of placement data by utilizing automated data collection and sophisticated NLP techniques.

Text summarization, a key component of this project, ad- dresses the challenge of processing large amounts of un- structured textual data [3]. While manual summarization can effectively capture main ideas, it requires considerable time and effort, making it impractical for large datasets. Automated text summarization (ATS) technologies can mitigate these issues by generating concise summaries that emphasize key insights while maintaining the original content's meaning and context [4]. This approach is particularly beneficial in placement scenarios, enabling students and administrators to quickly understand essential points from extensive textual data, such as feedback from seniors who have been placed. By employing ATS, users can significantly reduce the time and effort needed to extract relevant insights without having to read through lengthy documents.

Moreover, this project incorporates a chatbot as an inter- active interface that provides real-time responses to students' inquiries about placement information. Chatbots, which utilize NLP, are increasingly adopted across various sectors, including education, due to their ability to simulate conversations and deliver personalized information instantly [5], [6]. Implement- ing such a system in a placement context allows students to engage with the data more intuitively. The chatbot, trained on the information collected from placed seniors, will be capable of addressing a wide range of questions, offering a more accessible and user-friendly means for students to obtain placement-related information without manually sifting through numerous reports.

The evolution of chatbot technology, from early systems like ELIZA to contemporary virtual assistants such as Siri, Google Assistant, and IBM Watson, has enhanced these systems' intelligence and ability to manage complex tasks [7]. Modern chatbots are platform independent, require no installations,
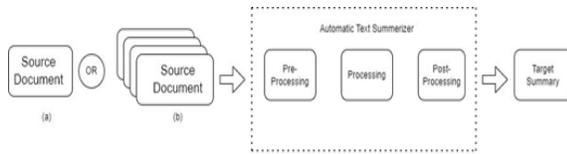
Figure 1. (a) Single-document or (b) Multi-document, automatic text summarizer.

Fig. 1. Automatic text summarizer

and provide instant access across multiple devices, making them particularly suitable for educational applications. These systems not only facilitate secure interactions but can also han- dle multiple conversations simultaneously, which is essential for addressing queries from numerous students about different companies. In this project, the chatbot and text summarizer will collaborate to create a seamless experience for students by delivering essential insights from placement data in real time. This integration of ATS and chatbot technologies will streamline data collection and summarization processes while enhancing the overall accessibility and usability of placement information. By automating these tasks, students will be empowered to make more informed decisions with minimal effort, ultimately improving their preparation for recruitment and placement activities. Through the incorporation of these advanced technologies, this project aims to transform the management of placement data, turning a traditionally labor- intensive task into an efficient and intelligent system. The chatbot's capability to provide real-time assistance, combined with the summarizer's ability to extract valuable insights from unstructured data, represents a significant advancement in how placement information can be processed and utilized by students, researchers, and administrators alike.

II. RELATED WORK

A summary typically condenses the original text to less than half its length, encapsulating the main ideas in a concise manner, often resulting in a paragraph-long summary. The journey of automatic document summarization began with Luhn [22], and since then, numerous methods have been devel- oped to identify key aspects of a text and create intermediate representations [8], [16], [17], [18], [19], [20], [21]. Early summarization techniques [22], [10] primarily re- lied on statistical methods, such as word frequency analysis, to pinpoint significant terms within documents. In contrast, contemporary approaches strive for a deeper semantic under- standing of the text. The summarization process is crucial for condensing content while preserving the essential meaning and informational features of the original text. Historically, manual summarization was labor-intensive and challenging; however, the automation of this process has become increasingly vital in academic research and studies [11]. Automatic text sum- marization (ATS) effectively reduces lengthy texts without sacrificing the main ideas and relevant information [12]. The goal of ATS is to distill large volumes of textual information into a more concise format, allowing readers to quickly grasp the main ideas.

ATS can be categorized into two primary types: a) Ex- tractive summarization, which identifies and extracts relevant passages from the original text using linguistic or statistical methods [13], and b) Abstractive summarization, which goes beyond merely repeating text, rephrasing it to create new expressions that may not appear in the original [14]. The significance of text summarization is evident in its diverse applications across various fields, including information re- trieval, news aggregation, document summarization, content creation, and clinical text summarization [12]. In particular, clinicians invest considerable effort in summarizing extensive textual information, such as compiling diagnostic reports and synthesizing patient treatment histories across multiple spe- cialists [15], [9].

Natural Language Processing (NLP) is essential for devel- oping intelligent chatbots, enabling them to comprehend and generate human-like responses. Recently, various NLP-based chatbot systems have emerged in fields such as education, re- cruitment, and customer support. For instance, Singh and Jain [23] created an NLP-driven chatbot utilizing lemmatization, named entity recognition (NER), and intent classification to assist students with university-related inquiries. Their work highlighted the importance of contextual understanding, al- though it primarily relied on static FAQ-style responses.

In the recruitment sector, Amin et al. [24] introduced a chatbot that employs NLP to match candidate resumes with job descriptions. While effective for initial screening, their system did not offer personalized responses based on previous user

experiences, which is a key feature of our approach. The advent of transformer models like BERT and GPT has significantly enhanced the capabilities of NLP-based chatbots by introducing contextual embeddings and attention mech- anisms [25]. Although these models achieve state-of-the-art performance, they often necessitate large datasets and sub- stantial computational resources. As an alternative, lightweight NLP techniques such as TF-IDF, sentence embeddings, and cosine similarity have proven effective for domain-specific applications [26].

In contrast to general-purpose chatbots, our system is specif- ically trained on summarized placement data gathered from senior students. It employs domain-specific preprocessing and sentence similarity techniques to provide relevant, real-time responses. This customized approach ensures a better align- ment with student inquiries regarding company roles, interview processes, and workplace culture, which are often overlooked in standard chatbot implementations.

III. PROPOSED METHODOLOGY

A. System Overview:

1) High-Level Architecture: The proposed system for man- aging placement data adopts a modular and user-centric archi- tecture, as depicted in Fig. 1. The frontend serves as the main access point for users, including students and administrators. Through this interface, users can engage with various modules, such as company information, a student dashboard, an admin dashboard, and a centralized portal for the chatbot and text
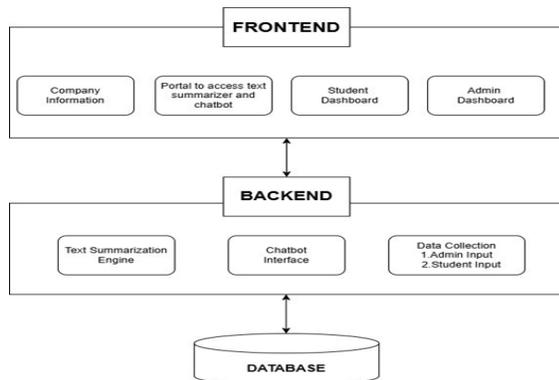

Fig. 2. System Architecture Diagram

summarizer. These components facilitate seamless data entry, retrieval of company records, and interaction with intelligent features like automated

summarization and conversational assistance. The backend supports these functionalities by or- chestrating requests and integrating three core subsystems: the text summarization engine, chatbot interface, and data col- lection module. This architecture promotes a clear separation of concerns, allowing for independent management of user interactions and backend logic, thereby enhancing scalability and maintainability.

2) Layered System Design: To illustrate the logical organi- zation of system components, a layered architectural model is presented in Fig. 3. This model divides the system into five distinct layers. At the top, the User Interface Layer enables interactions from end-users, namely students and administra- tors. Below this is the Authentication Layer, which manages user credentials through secure registration, login, and logout processes. The Application Layer contains the core functional logic of the system, encompassing both frontend and backend components that handle user requests and backend services. These services are further supported by the Functionality Layer, which includes the chatbot, text summarizer, and information collection modules. Additionally, an Integration Layer facilitates interaction with external APIs and services, extending the system's capabilities. At the base, the Data Layer ensures persistent storage of all structured information using a secure database management system (DBMS), enhancing confidentiality and robustness. This layered design improves modularity, security, and the potential for future expansion.

B. Text Summarization Method:

1) Feature Extraction: The feature extraction process utilizes TF-IDF vectorization to transform textual recruitment data into numerical representations. This method evaluates term frequency while considering document frequency, effectively identifying and weighting domain-specific keywords. By high- lighting distinctive terms and minimizing common words, it generates meaningful numerical features that capture essential recruitment patterns while remaining computationally efficient for analysis. This transformation retains key semantic informa- tion, enabling machine processing of the textual data.

2) Sentence Ranking: The system ranks sentences using a graph-based PageRank approach.

Initially, it computes pair- wise cosine similarities between sentence vectors to create a similarity matrix. These similarity scores form a weighted graph where sentences are nodes and similarities are edges. The PageRank algorithm analyzes this graph structure to identify the most central and significant sentences based on their connections, effectively highlighting key content while preserving semantic relationships. This method provides an efficient means to assess sentence importance within the document's context.

3) Extractive Summarization: The extractive summarization selects the top sentences based on PageRank scores while maintaining their original order for coherence, with fallback handling for edge cases. This output is then processed in a hybrid enhancement stage where generative AI (Gemini) uses the extractive summary as context to generate deeper insights while minimizing hallucinations. This combined approach delivers both factual accuracy from extraction and analyti- cal depth from generation, making it ideal for processing recruitment data where precision and interpretation are equally important.

C. Chatbot Design:

1) Type of Chatbot: The chatbot is developed as a rule- guided, generative system integrated into this project. It is designed to address placement-related queries and leverages Google's Gemini-1.5-Flash model to generate accurate re- sponses based on current data. The implementation utilizes the Django web framework, functioning as an interactive web-based service accessible to both students and administrators.

2) Use of Summarized Data as Knowledge Base: To ensure relevance and accuracy, the chatbot relies on a database con- taining placement data. This data, which may be preprocessed through summarization techniques, forms the core knowledge base. For each user query, the chatbot retrieves this data in real-time, grounding its responses in the most up-to-date information available.

3) Components: Input Processing and Response Matching: The architecture consists of two layers: frontend and backend. The frontend (HTML interface) collects user input, which is sent to the backend via HTTP POST requests. The backend processes the input and constructs a context-driven prompt that combines the placement data with the user's question. If the answer cannot be derived from the provided data, the model
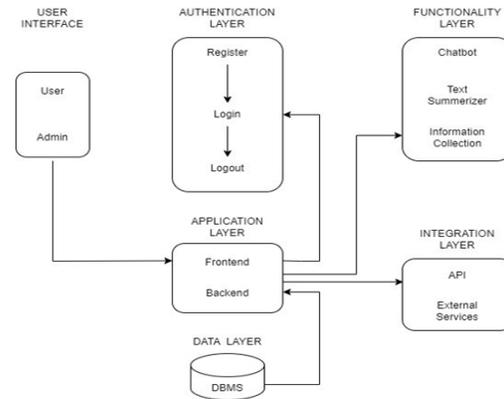

Fig. 3. Layered System Architecture Diagram

is instructed to respond with" Your question is out of scope." This structured prompt is then sent to the Gemini model via the Generative AI API, and the output is returned to the frontend as a JSON response, enabling real-time interaction. Robust error handling ensures graceful fallback in the event of network or data retrieval failures.

4) Development and Integration: The chatbot integrates the Gemini-1.5-Flash model through the Google Generative AI API. It utilizes real-time placement data from a database, which is combined with user queries to create structured prompts. These prompts are sent to the model using the generate_ content () method, and the responses are returned in JSON format. Although the model's internals are abstracted, Gemini is based on a transformer architecture with standard components such as tokenized input, hidden layers with ReLU activation, and output generated via Softmax. The model is pre-trained using categorical cross-entropy loss and does not require additional training or fine-tuning. Input encoding and response generation are managed internally by the API, making the chatbot lightweight and easy to integrate into a web application.

IV. CHALLENGES

A. Data Quality and Completeness:

A significant challenge in this project is ensuring that the data collected from students is both accurate and comprehen- sive. Given that the system relies on

minimal input—such as email addresses and company names—there is a risk of omitting crucial details or entering incorrect information. Stu- dents may also provide inconsistent or incomplete data, which can adversely affect overall data quality. While automating data collection through forms simplifies the process, it also increases the likelihood of errors or omissions. Ensuring
the completeness and reliability of this data is essential for generating meaningful insights and accurate responses from the chatbot.

B. Text Summarization Accuracy:

Another challenge involves developing a text summarizer that consistently produces useful and accurate summaries from the unstructured data provided by students. Recruitment experiences and company feedback can be articulated in various ways, depending on individual students' expressions. These differences in language, structure, and detail complicate the summarizer's task of extracting the most relevant points while preserving context. Consequently, there is a risk of generating summaries that either overlook important insights or become overly generalized. Striking a balance between conciseness and comprehensiveness is crucial for ensuring the summarizer's effectiveness.

C. Training the Chatbot Effectively:
Training the chatbot to accurately respond to students' inquiries about specific companies is a complex task. It necessitates well-organized, detailed, and relevant data for each company, which must be structured carefully for the chatbot to understand and respond appropriately. The chatbot must also accommodate variations in how students phrase their questions, which can range from simple, direct queries to more complex, nuanced ones. Ensuring that the chatbot interprets these questions correctly and provides accurate, helpful responses is vital for the project's success.

D. Handling Diverse Query Types:
Students may ask a wide array of questions, some of which may fall outside the system's data scope or may be poorly phrased. Designing the chatbot to effectively

manage such diverse queries presents a significant challenge. The chatbot must differentiate between relevant and irrelevant informa- tion, clarify ambiguous questions, and provide meaningful responses. Additionally, it must handle queries related to various topics, from general company overviews to specific placement processes. Ensuring that the chatbot delivers accu- rate responses across this spectrum will require sophisticated natural language processing and machine learning techniques.

E. Data Privacy and Security:
Given that the system processes personal data, such as students' email addresses and details about their placement experiences, ensuring the privacy and security of this infor- mation is paramount. The risk of data breaches, unauthorized access, or misuse of personal information is a major concern. Compliance with data protection regulations (such as GDPR or local privacy laws) must be a top priority to safeguard students' information and maintain their trust in the system. Implementing robust security measures, including encryption and secure access controls, will be crucial for protecting sensitive data.

F. Customization and Personalization:
Providing personalized responses through the chatbot that cater to individual students' specific queries and preferences is a challenging task. This requires the system to process and un- derstand each student's unique context, background, and query history. The complexity of implementing such customization lies in accurately identifying and matching students' needs with relevant data, which can vary significantly between users. Effective natural language processing algorithms are essential for delivering personalized responses, making this a critical and challenging aspect of the project.

G. Real-time Processing:
Another challenge is ensuring that the system can process and summarize the collected data in real time while main- taining accuracy and relevance. As data is gathered from multiple students and potentially updated frequently, the sys- tem must quickly analyze and summarize this information without compromising output quality. The chatbot also needs to provide real-time responses to users' queries, which adds to the processing demands. Efficiently handling

large volumes of data and delivering timely responses will require careful optimization of both the text summarizer and the chatbot.

V. CONCLUSION

This project addresses the growing complexity of managing and accessing placement data for students by integrating AI technologies such as text summarization and chatbots. The automated collection and processing of data streamline information flow, making it more accessible and useful for students preparing for placements. By tackling issues like the limited application of AI in placement systems, challenges in processing unstructured data, and the lack of real-time, personalized assistance, this solution offers a modern approach to addressing inefficiencies in placement preparation. The use of advanced natural language processing techniques ensures that data is summarized and presented meaningfully, allowing students to make informed decisions based on concise, action-able insights. Additionally, the chatbot provides an interactive platform to help students quickly and easily find relevant information, enhancing their overall placement experience. This system has the potential to significantly improve how placement data is managed and utilized in educational institu- tions, laying the groundwork for further AI-driven innovations in student support systems.

REFERENCES

[1] Gupta and P. S. Saini," A review on natural language processing (NLP): Applications, challenges, and research directions," Materials Today: Proceedings, vol. 60, pp. 394-400, 2022.

[2] M. A. Hearst," Untangling Text Data Mining." Proceedings of ACL, 1999. C.-Y. Lin and E. Hovy," Automatic evaluation of summaries using n-gram co-occurrence statistics," in Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics, 2003.

[3] C.-Y. Lin and E. Hovy," Automatic evaluation of summaries using n- gram co-occurrence statistics," in Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics, 2003.

[4] R. Nallapati, B. Zhou, C. Gulcehre, and B. Xiang." Abstractive Text Summarization using Sequence-to-Sequence RNNs and Beyond," arXiv preprint arXiv: 1602.06023, 2016.

[5] S. Hussain, O. A. Kamal, and A. N. Qamar," The use of chatbots in education: A systematic literature review," IEEE Access, vol. 9. pp. 162361-162379, 2021.

[6] Y. Fadhil," C a n a chatbot support mental health and well-being in the workplace?" in Proceedings of the 32nd International BCS Human Computer Interaction Conference (HCI 2018), 2018.

[7] M. McTear," T h e rise of the conversational interface: A new kid on the block?" in Proceedings of the 2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS). 2016, pp. 506-511.

[8] Suleiman and A. Awajan," Deep learning-based abstractive text summarization: Approaches, datasets, evaluation measures, and chal- lenges," Math. Problems Eng., vol. 2020, no. 1, 2020, Art. no. 9365340.

[9] S. L. Fleming et al.," Medalign: A clinician-generated dataset for instruction following with electronic medical records," in Proc. AAAI Conf. Artif. Intell., 2024 pp. 22021-22030.

[10] S. Masoumi, M. Feizi-Derakhshi, and R. Tabatabaei," Tabsum-a new Persian text summarizer," J. Math. Comput. Sci., vol. 11, no. 4, pp. 330-342, Aug. 2014.

[11] N. Alami, M. Meknassi, N. En-nahnahi, Y. El Adlouni, and O. Ammor," U n s u p e r v i s e d neural networks for automatic Arabic text summarization using document clustering and topic modeling." Expert Syst. Appl., vol. 172, May 2021, Art. no. 114652, doi: 10.1016/j.eswa.2021.114652.

[12] Khan, Z. A. Shah, M. Usman, I. Khan, and B. Niazi," Exploring the landscape of automatic text summarization: A comprehensive survey," IEEE Access, vol. 11, pp. 109819-109840, 2023, doi: 10.1109/AC- CESS.2023.3322188.

[13] Suleiman and A. Awajan," Deep learning based abstractive text summarization: Approaches, datasets, evaluation measures, and chal- lenges," Math. Problems Eng., vol. 2020, no. 1, Aug. 2020,

Art. no. 9365340.

[14] Z. Liang, J. Du, and C. Li," Abstractive social media text summarization using selective reinforced Seq2Seq attention model," Neurocomputing. vol. 410, pp. 432-440, Oct. 2020.

[15] B. G. Arndt, J. W. Beasley, M. D. Watkinson, J. L. Temte, W.-J. Tuan,Sinsky, and V. J. Gilchrist," Tethered to the EHR: Primary care physician workload assessment using EHR event log data and time- motion observations," Ann. Family Med., vol. 15, no. 5, pp. 419-426. Sep. 2017.

[16] Guo, R. Pasunuru, and M. Bansal," Soft layer-specific multi-task sum-marization with entailment and question generation," in Proc. 56th Annu. Meeting Assoc. Comput. Linguistics, Melbourne, VIC, Australia, 2018, pp. 687-697.

[17] S. Subramanian, A. Trischler, Y. Bengio, and C. J. Pal," Learning general purpose distributed sentence representations via large-scale multi-task learning," in Proc. 6th Int. Conf. Learn. Represent. (ICLR), Vancouver, BC, Canada, 2018, pp. 1-11.

[18] Devlin, M. Chang, K. Lee, and K. Toutanova," BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol. (NAACI-HLT), Minneapolis, MN, USA, Jan. 2018, pp. 4171- 4186.

[19] R. Kiros, Y. Zhu, R. Salakhutdinov, R. S. Zemel, A. Torralba, R. Urtasun, and S. Fidler," Skip-thought vectors," Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 28, pp. 3294-3302, Dec. 2015.

[20] M. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer," Deep contextualized word representations," in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol., vol. 1, New Orleans, LA, USA, 2018, pp. 2227-2237.

[21] Joshi, E. Fidalgo, E. Alegre, and L. Ferna´ndez-Robles," Summ- Coder: An unsupervised framework for extractive text summarization based on deep auto-encoders," Expert Systems with Applications, vol. 129, pp. 200–215, Sep. 2019.

[22] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed," Au- tomatic text summarization: A comprehensive survey," Expert Systems with Applications, vol. 165, Mar. 2021, Art. no. 113679.

[23] R. Singh and A. Jain," SmartBot: An NLP-based chatbot for university admission system," Int. J. Adv. Comput. Sci. Appl. (IJACSA), vol. 12, no. 3, pp. 123–129, 2021.

[24] Amin, M. A. Nawaz, and M. Waqas," AI-based Recruitment Chatbot: A Case Study of Applicant Screening," in Proc. Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST), 2019.

[25] Devlin, M. W. Chang, K. Lee, and K. Toutanova," BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, 2019.

[26] N. Reimers and I. Gurevych," Sentence-BERT: Sentence embeddings using Siamese BERT-networks," in Proc. EMNLP-IJCNLP, 2019