

# Enhancing International Graduate Student Experience through AI-Driven Support Systems: A LLM and RAG-Based Approach

1<sup>st</sup> Binita Saha  
Dept. Of Computer Science  
North Dakota State University  
Fargo, ND, USA  
binita.saha@ndsu.edu

2<sup>nd</sup> Utsha Saha  
Dept. Of Computer Science  
North Dakota State University  
Fargo, ND, USA  
utsha.saha@ndsu.edu

**Abstract**—International graduate students encounter unique challenges that impede their academic and personal success. This paper introduces an AI-powered chatbot designed specifically for these students, utilizing advanced language models and Retrieval-Augmented Generation (RAG). Unlike generic solutions, our chatbot is tailored with a dataset curated from Reddit communities frequented by international students, enabling it to provide highly relevant and actionable advice. The system combines GPT-3.5's generative capabilities with precise information retrieval to effectively guide students through academic procedures, cultural adjustments, and personal challenges. An evaluation shows that our RAG-enhanced model outperforms standard GPT-3.5, demonstrating significant improvements in response accuracy and relevance. This research not only advances AI applications in student support but also offers practical, real-time aid to enhance international students' educational experiences.

**Keywords**—International students, Natural language processing (NLP), GPT-3.5-turbo, Retrieval-Augmented Generation (RAG), Personalized support systems, Chatbot

## I. INTRODUCTION

In the dynamic world of global academia, international graduate students are invaluable, bringing diverse perspectives that enrich the academic community and foster global collaboration. However, these students frequently face a variety of complex challenges that hinder both their academic progress and personal adjustment. These challenges are not limited to academic issues but include linguistic barriers, cultural adaptation difficulties, and logistical challenges such as finding accommodation and adapting to unfamiliar educational systems. Research indicates that these obstacles often require solutions that go beyond the capabilities of traditional institutional support systems [1] [2]

Existing support frameworks often fail to meet the immediate and specific needs of international students, leading to significant disparities in their academic experiences and outcomes. Addressing this deficiency, our study introduces an innovative application of cutting-edge natural language processing technologies. We employ the GPT-3.5-turbo model, enhanced with Retrieval-Augmented Generation (RAG), crafted specifically for the international student demographic. This approach is unique; while RAG has been applied in various fields, our implementation targets the precise needs of international students by utilizing a dataset specifically curated from discussions and

interactions within subreddit communities dedicated to these students.

This research presents a specialized adaptation of the GPT-3.5-turbo model with RAG to develop a chatbot that provides tailored, real-time support. This system leverages community-driven data to offer solutions that are not only relevant and timely but also highly personalized, thus addressing the unique challenges faced by this group. By analyzing and integrating insights from actual student interactions, the chatbot delivers support that is both practical and aligned with the students' specific circumstances, significantly enhancing the responsiveness and effectiveness of support mechanisms.

The development and application of this AI-driven chatbot are detailed in this paper, demonstrating its superior performance over traditional models in delivering relevant and personalized support. This technology sets a new standard in the application of AI for enhancing the support systems for international students.

Following this introduction, the paper is structured to first review the literature concerning the challenges faced by international students and the current mechanisms in place to support them. It then describes the methodology employed in customizing the RAG system specifically for this demographic. Subsequent sections present a comparative analysis of the chatbot's performance and discuss our findings and future research directions.

## II. RELATED WORKS

International students face numerous challenges while pursuing education in foreign countries, significantly impacting their educational experience and overall well-being. These challenges primarily revolve around academic, linguistic, cultural, and accommodation-related difficulties [3]. Differences in educational systems, curriculum structures, and teaching methodologies between home countries and host institutions can lead to academic difficulties, such as navigating unfamiliar educational systems, adapting to new coursework requirements, striving to meet distinctive academic objectives, developing critical thinking skills, and mastering academic writing in a non-native language [4]–[7].

Cultural adjustment and language barriers also pose significant challenges for international students. Fitting into a new culture can be demanding, and language barriers can hinder effective communication, academic engagement, and social integration [8]–[11]. Cultural differences, including social norms, attitudes, behaviors, and environmental factors, can sometimes lead international students to feel disoriented and isolated, significantly impacting their overall academic experiences and social integration [4], [9]–[11]. As these challenges accumulate, they can contribute to increased stress and depression levels among international students, with homesickness, cultural shock, and discrimination being the leading stress-causing factors [12].

Although existing support mechanisms, such as international student offices, peer mentorship programs, and language courses, provide valuable assistance, they may not always deliver timely and precise guidance, resulting in inconsistent support. Recent advancements in artificial intelligence, particularly chatbots, have emerged as a promising technology to assist students in various aspects of their academic journey.

Choque-Díaz et al. [13] proposed a cognitive technology model leveraging chatbots to enhance academic support services for international graduate students, emphasizing the need for personalized support beyond academic struggles. Ramandanis and Xinogalos [2] conducted a systematic literature review investigating the support provided by chatbots to educational institutions and their students, suggesting that a Retrieval-Augmented Generation (RAG) model can be effectively used to develop chatbots that offer personalized support addressing academic, cultural, and personal challenges.

The integration of chatbots on social media platforms has also been explored to facilitate language learning. Sarosa et al. [14] developed a social media-based chatbot for English learning, leveraging users' familiarity with the platform's interface and navigation. Huang [15] designed and evaluated chatbot-enhanced activities for a flipped graduate course, demonstrating the potential of chatbots as pre-class activities. Höhn [15] proposed a data-driven model of explanations for a chatbot that helps practice conversation in a foreign language, validated in an AIML-based chatbot, showcasing its effectiveness in language learning scenarios. Colace et al. [1] developed a chatbot system to provide support to university students on specific courses, highlighting the potential of chatbots in e-learning contexts.

The application of deep learning techniques in chatbot development has garnered significant attention. Bhartiya et al. [16] developed an artificial neural network-based university chatbot system, improving the probability scores of correct responses through iterative training. Kulkarni et al. [17] explored the use of reinforcement learning for optimizing RAG in domain-specific chatbots, showcasing the potential for further optimization in chatbot performance.

Despite these developments, there remains a gap in providing real-time, tailored support that addresses the specific and immediate needs of international students. Our research aims to fill this gap by enhancing the personalization and responsiveness of chatbot interactions using a Retrieval-Augmented Generation model, building a more effective and supportive educational

environment for international students.

### III. METHODOLOGY

The objective of this research is to harness the power of LLM to enhance the support provided to international graduate students. To achieve this, a systematic methodology (Fig. 1) was drawn up and is detailed below:

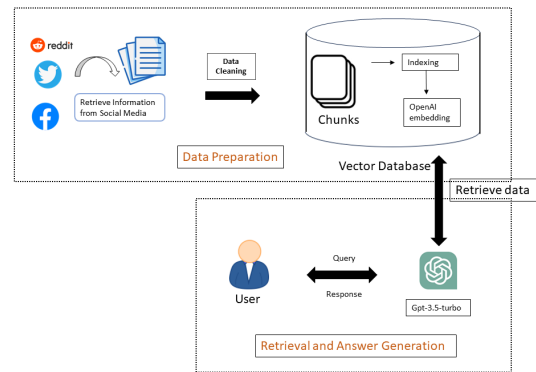


Fig. 1. System architecture of Overall RAG system

#### A. Data Collection and Preprocessing

To build a dataset that fits our research goals, we mainly used Reddit as our main data source. Reddit is known for its rich community-driven discussions and user-generated content. However, our system's flexible setup allows for easy integration with other social media platforms, like Twitter or Facebook.

When gathering data, we carefully picked three well-known subreddits based on their relevance and popularity: '/phd' with a large following of 106k members, '/international student' with 21k users, and '/inttousa' with 7.2k dedicated participants. We collected the top 1000 posts and their comments from each subreddit. After gathering information from these subreddits, the dataset was processed for chatbot training. To keep the posts authentic and diverse, we combined posts with similar themes using semantic embeddings. This technique helped identify and group related posts, and their comments were combined into a single line for each post.

Keeping the dataset rich was crucial, but we also maintained the highest privacy standards. We systematically went through the data, making sure to remove any Personally Identifiable Information (PII) to protect the anonymity and privacy of the contributors. At the same time, unwanted content like ads, spam, and off-topic discussions were removed during the data cleaning and preprocessing stage.

#### B. Bias Detection and Data Categorization

To ensure bias detection and adherence to ethical standards, we integrated advanced natural language processing techniques and machine learning algorithms. These methods identified and flagged potentially biased content, such as discriminatory language or offensive remarks. A synonym dictionary expanded the coverage of problematic terms [18]. Flagged content was reviewed by experts and removed if found biased or unethical.

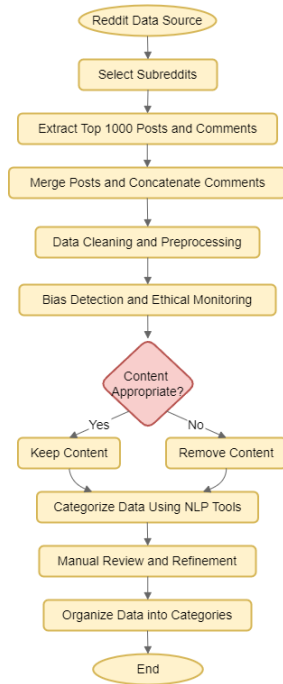


Fig. 2. Data Collection and Preprocessing

Ongoing monitoring and audits ensured the chatbot remains unbiased and ethically sound.

Post bias detection, the data was categorized using advanced language processing tools into five themes: Admission Process and Application Procedures, University Selection, Recommendation Letters, Academic Stress and Mental Health, and Language Barrier and Cultural Adaptation. A manual review of a subset validated the categorization accuracy.

By incorporating these mechanisms throughout data collection and preprocessing, we ensured our dataset is comprehensive, relevant, unbiased, and ethically sound. This guarantees that the chatbot will provide accurate, fair, and trustworthy information to international students seeking guidance and support.

### C. Configuring the GPT-3.5 Model

The next phase involves configuring the GPT-3.5 model and integrating it with RAG using the preprocessed dataset. This process begins with setting up and configuring the environment for GPT-3.5-turbo, ensuring that all necessary libraries and dependencies are in place. Following this, the GPT-3.5-turbo is connected with a retrieval system capable of accessing the knowledge base created from social platforms. The preprocessed dataset is then utilized to identify key phrases and topics that will guide the retrieval system in fetching relevant information from external sources.

### D. Retrieval-Augmented Generation (RAG) Implementation

To implement the RAG component effectively in our chatbot system, the retrieval mechanism efficiently finds the best match for user questions from our extensive knowledge base. This setup involved several key steps:

1) *Vector Embedding*: Each piece of data was transformed into embedding vectors using OpenAI's "text-embedding-3-small" model, ensuring efficient indexing and retrieval based on semantic similarity.

2) *Query Processing*: User queries are converted into vectors using the same embedding model, allowing for semantic searches against our pre-indexed dataset.

3) *Semantic Search and Data Retrieval*: The system searches the indexed vector database to identify and retrieve the most relevant posts and comments matching the semantic content of the user's query. We set the retrieval system to consider the top two closest matches ( $k=2$ ) for well-informed and contextually appropriate responses.

4) *Response Generation*: The GPT-3.5 model uses the retrieved data to generate responses that are accurate and tailored to the specific context of the user's inquiry, ensuring relevance and usefulness.

5) *Handling Out-of-Scope Queries*: For queries outside the available data, the system recognizes these limitations and informs the user accordingly, avoiding potentially inaccurate or misleading information.

**Context**: An international student is struggling with English language skills, which affects their ability to participate in class discussions and social activities, leading to feelings of isolation

**User**: I'm having trouble following lectures and joining discussions because my English isn't very good. What can I do to improve?

**GPT 3.5 without RAG**: Improving your language skills takes time. Try to practice as much as you can by reading, writing, and speaking in English. Watching English movies and listening to podcasts can also be helpful. Consider joining a language class if you haven't already.

**GPT 3.5 with RAG**: I understand how challenging language barriers can be. For academic lectures, you might want to use lecture recording tools, which some universities offer, so you can replay them at your own pace. This can help you catch details you might miss the first time. For improving conversational English, consider joining a 'Conversation Buddy' program, which pairs you with a fluent English speaker. This program is specifically designed to help international students improve practical speaking skills in a low-pressure environment. Also, your university might offer specific workshops on academic English that are tailored for non-native speakers which could be incredibly beneficial.

Fig. 3. Comparison of Standard LLM and RAG-Enhanced-LLM responses to International Student Language Challenge

Fig. 3 shows that the custom RAG-LLM provides specific, actionable advice addressing both academic and social aspects of the student's language barrier. By suggesting tailored university resources and programs, the response is directly relevant and useful, whereas the regular LLM provides general information lacking specific guidance.

This comprehensive methodology ensures accurate responses to user queries. Given that large language models (LLMs) can sometimes generate fabricated responses or hallucinate, the use of a prompting mechanism mitigates these issues and enhances the reliability of the responses.

### E. Testing with RAGAS and BERTScore

The evaluation of the chatbot system designed to support international graduate students utilized two key metrics: Retrieval-Augmented Generation Answer Similarity (RAGAS) [19] and BERTScore [20]. RAGAS measures the faithfulness, relevance, context precision, and context recall of responses against reference answers, assessing the accuracy of the RAG-enhanced chatbot. BERTScore uses pre-trained BERT models

to compare the semantic similarity between generated responses and reference answers, computing precision, recall, and F1 scores.

The testing process involved generating responses for diverse user queries by both the baseline GPT-3.5-turbo model and the RAG-enhanced model. A set of reference answers were curated for the testing. RAGAS evaluated the faithfulness, relevance, context precision, and context recall of the responses, while BERTScore assessed their semantic similarity. The following formulas are used to calculate precision, recall, and F1 score:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

The results from both models were compared to determine the improvements achieved by the RAG-enhanced model. The structured testing methodology using RAGAS and BERTScore provided a comprehensive evaluation of the chatbot's performance.

#### IV. RESULTS AND DISCUSSION

This section presents the experimental results and a comprehensive analysis of the RAG-enhanced LLM's performance. We evaluate our model based on various metrics (Table I), including RAGAS [19] and BERTScore [20], and discuss its effectiveness in providing tailored support to international graduate students.

TABLE I

COMPARISON OF EVALUATION METRICS WITH AND WITHOUT RAG

| Framework | Evaluation Matrix | without RAG | with RAG |
|-----------|-------------------|-------------|----------|
| RAGAS     | Faithfulness      | 0.56        | 0.92     |
|           | Answer Relevancy  | 0.65        | 0.95     |
|           | Context Precision | 0.43        | 0.82     |
|           | Context Recall    | 0.38        | 0.71     |
|           | Harmfulness       | 0           | 0        |
| BERTScore | Precision         | 0.22        | 0.59     |
|           | Recall            | 0.32        | 0.66     |
|           | F1 Score          | 0.27        | 0.63     |

The evaluation of the model's performance reveals significant improvements with the incorporation of RAG. The metrics demonstrate that the RAG-enhanced GPT-3.5 model consistently outperforms the base model without RAG across all evaluation criteria. Table I presents the data in tabular format, while Fig. 4 illustrates it graphically.

##### A. RAGAS Faithfulness and Answer Relevancy

The faithfulness metric improved from 0.56 to 0.92 with the inclusion of RAG, indicating that the responses generated with RAG are more accurate and reliable. Similarly, the answer relevancy increased from 0.65 to 0.95, demonstrating that the responses are more pertinent and closely aligned with the user queries. These improvements underscore the effectiveness of RAG in enhancing the quality and reliability of the generated content.

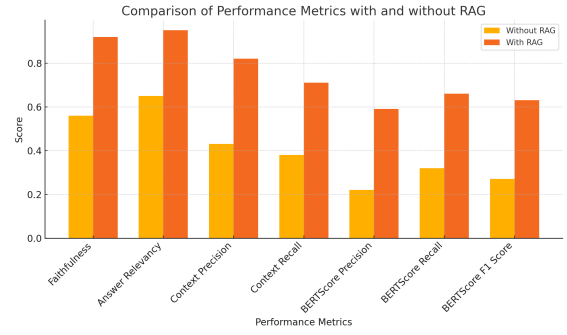


Fig. 4. Comparison Of Performance Metrics With And Without RAG

##### B. RAGAS Context Precision and Recall

The context precision and recall metrics experienced substantial improvements. Context precision increased from 0.43 to 0.82, and context recall rose from 0.38 to 0.71. These metrics demonstrate that the RAG-enhanced model excels at maintaining and understanding the context of user queries, resulting in more coherent and contextually appropriate responses.

##### C. RAGAS Harmfulness

Both systems scored zero for harmfulness, reflecting that neither generated harmful or inappropriate content. This outcome underscores the importance of the preprocessing steps taken to ensure data quality and the ethical considerations embedded in the system's design.

##### D. BERTScore Precision and Recall

The BERTScore precision and recall metrics further highlight the improved performance. The BERTScore precision jumped from 0.22 to 0.59, while the recall improved from 0.32 to 0.66. These scores reflect the model's enhanced capability to generate responses that are semantically similar to the expected answers, validating the relevance and appropriateness of the generated content.

##### E. BERTScore F1 Score

The overall F1 score is the mean value of precision and recall, improved from 0.26 to 0.62. This significant increase indicates a well-rounded enhancement in the model's performance, with the RAG integration contributing to more accurate, and reliable outputs.

##### F. Discussion

The results clearly demonstrate the advantages of integrating RAG with gpt-3.5-turbo for domain specific information retrieval. The RAG-enhanced model not only provides more accurate and relevant responses but also better understands and maintains the context of user queries. These improvements are crucial for offering personalized and efficient support to international graduate students.

## V. CONCLUSION AND FUTURE WORK

This research explored integrating Retrieval-Augmented Generation (RAG) with GPT-3.5-turbo to support international graduate students. By combining GPT-3.5's generative capabilities with targeted information retrieval, we aimed to address the unique challenges faced by this student population. Our methodology included data collection from Reddit, preprocessing to ensure privacy and relevance, and implementing RAG to enrich the model's responses.

The experimental results, evaluated using metrics such as RAGAS [19] and BERTScore [20], demonstrated that the RAG-enhanced model can provide more contextually relevant and accurate support compared to the base GPT-3.5 model. However, there are several areas for improvement.

A primary limitation is the reliance on a curated knowledge base, which restricts the model's ability to provide real-time updates. Future work should focus on integrating dynamic data sources to enhance responsiveness. Additionally, incorporating more comprehensive human evaluations would offer deeper insights into the quality and usability of the responses.

Another significant concern is the potential for biased or inappropriate responses. This is crucial and should be emphasized more in the preprocessing stage to reduce misinformation and potential biases in the dataset. These aspects are particularly important when handling sensitive information from platforms like Reddit, which are often unmoderated. Future work should include robust mechanisms for bias detection and continuous monitoring to ensure ethical standards. Expanding the knowledge base to include more diverse sources and perspectives can also improve the model's ability to address a wider range of queries.

Moreover, exploring additional AI techniques, such as reinforcement learning and advanced NLP methods, could refine the model further. Implementing user feedback loops and adaptive learning algorithms can help the system evolve and better meet the needs of international graduate students. Details on how user feedback will be integrated into the system to continually improve the chatbot's responses should be added.

While the chatbot aims to offer personalized guidance, its ability to fully understand the nuanced and context-specific needs of individual students is limited. Personalization is constrained by the data and the model's capacity to interpret complex, context-dependent issues. The dataset used comprises Reddit posts and comments, which may not fully represent the diverse experiences and challenges of international graduate students. This data might be biased towards individuals who are more likely to seek advice online, potentially overlooking the experiences of those who do not use Reddit or other online forums. Future work should discuss these limitations and explore alternate approaches, including different RAG frameworks, to improve the quality of the paper.

In conclusion, while there are limitations, this research demonstrates the potential of RAG-enhanced GPT-3.5-turbo to offer personalized support for international graduate students. Addressing these challenges and building upon the current framework can lead to more robust and effective solutions in AI-driven student support systems.

## REFERENCES

- [1] F. Colace, M. D. Santo, M. Lombardi, F. Pascale, A. Pietrosanto, and S. Lemma, "Chatbot for e-learning: A case of study," *International Journal of Mechanical Engineering and Robotics Research*, vol. 7, pp. 528–533, 2018.
- [2] D. Ramandanis and S. Xinogalos, "Investigating the support provided by chatbots to educational institutions and their students: A systematic literature review," *Multimodal Technologies and Interaction*, vol. 7, no. 11, p. 103, 2023.
- [3] B. Talebloo and R. B. Baki, "Challenges faced by international postgraduate students during their first year of studies," *International Journal of Humanities and Social Science*, vol. 3, no. 13, pp. 138–145, 2013.
- [4] A. G. Trice, "Faculty perceptions of graduate international students: The benefits and challenges," *Journal of studies in International Education*, vol. 7, no. 4, pp. 379–403, 2003.
- [5] G. Harman, "International phd students in australian universities: Financial support, course experience and career plans," *International Journal of Educational Development*, vol. 23, no. 3, pp. 339–351, 2003.
- [6] S. Pappa, M. Elomaa, and S. Perälä-Littunen, "Sources of stress and scholarly identity: the case of international doctoral students of education in finland," *Higher Education*, vol. 80, no. 1, pp. 173–192, 2020.
- [7] A. B. Click, "International graduate students in the united states: Research processes and challenges," *Library & Information Science Research*, vol. 40, no. 2, pp. 153–162, 2018.
- [8] A. A. A. Alghail and O. H. A. Mahfoodh, "Academic reading difficulties encountered by international graduate students in a malaysian university," *Issues in Educational Research*, vol. 26, no. 3, pp. 369–386, 2016.
- [9] V. Womujuni, "The challenges international students face in adjusting to their new status as graduate students: An exploratory case study," PhD thesis, Portland State University, January 2006, available at <https://doi.org/10.15760/etd.5870>.
- [10] Y.-H. Kuo, "Language challenges faced by international graduate students in the united states," *Journal of International Students*, vol. 1, no. 2, 2011.
- [11] T. M. Karkar-Esperat, "International graduate students' challenges and learning experiences in online classes," *Journal of International Students*, vol. 8, no. 4, pp. 1722–1735, 2018.
- [12] W. T. Gebregergis, "Major causes of acculturative stress and their relations with sociodemographic factors and depression among international students," *Open Journal of Social Sciences*, vol. 6, no. 10, pp. 68–87, 2018.
- [13] M. Choque-Díaz, J. Armas-Aguirre, and P. Shiguihara-Juárez, "Cognitive technology model to enhanced academic support services with chatbots," in *2018 IEEE XXV International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*, 2018, pp. 1–4.
- [14] M. Sarosa, M. Kusumawardani, A. Suyono, and M. H. Wijaya, "Developing a social media-based chatbot for english learning," in *IOP Conference Series: Materials Science and Engineering*, vol. 732, 2020, p. 012074.
- [15] S. Höhn, "A data-driven model of explanations for a chatbot that helps to practice conversation in a foreign language," in *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, 2017, pp. 395–405.
- [16] N. Bhartiya, N. Jangid, S. Jannu, P. Shukla, and R. Chapaneri, "Artificial neural network based university chatbot system," in *2019 IEEE Bombay Section Signature Conference (IBSSC)*, 2019, pp. 1–6.
- [17] M. Kulkarni, P. Tangarajan, K. Kim, and A. Trivedi, "Reinforcement learning for optimizing rag for domain chatbots," *arXiv preprint arXiv:2401.06800*, 2024.
- [18] M. I. Hossain, U. Saha, and A. Billa, "The application of a synonyms dictionary using bi-lstm in bangla for better approximation," in *2023 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, 2023, pp. 234–238.
- [19] S. Es, J. James, L. Espinosa-Anke, and S. Schockaert, "Ragas: Automated evaluation of retrieval augmented generation," 2023.
- [20] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, "Bertscore: Evaluating text generation with bert," 2020.