Cilt/Volume 15 | Sayı/Issue 1 | Haziran/June 2026 | 272-287

Digital Student Services: Intelligent Chatbot Systems Supported by Artificial Intelligence and RAG Model

Muhammed Fatih ÖZER*, Musab Talha AKPINAR**, Semih CEYHAN***

ABSTRACT

Purpose: This study develops and evaluates a RAG-based AI chatbot for university student services.

Methodology/Approach: Preliminary interviews with 20 students informed the design, and post-deployment evaluation with 40 students assessed response quality, reliability and usability.

Findings: The chatbot improved accessibility, response efficiency and user satisfaction by retrieving current institutional information.

Practical implications: Integration into student affairs portals and mobile applications can reduce administrative workload and support continuous student access.

Originality: The study demonstrates the use of retrieval-augmented generation for context-aware, accurate and scalable digital student services.

Keywords: Artificial Intelligence, Large Language Models, Retrieval-Augmented Generation, Natural Language Processing, Digital Student Services

JEL Codes: I23, O33

Dijital Öğrenci Hizmetleri: Yapay Zekâ ve RAG Modeli Destekli Akıllı Sohbet Robotu Sistemleri

ÖZ

Amaç: Bu çalışma, üniversite öğrenci hizmetleri için RAG tabanlı bir yapay zekâ sohbet robotu geliştirmeyi ve değerlendirmeyi amaçlamaktadır.

Yöntem: Tasarım süreci 20 öğrenciyle yapılan ön görüşmelerle şekillendirilmiş; uygulama sonrası 40 öğrenciyle yanıt kalitesi, güvenilirlik ve kullanılabilirlik değerlendirilmiştir.

Bulgular: Sohbet robotu, güncel kurumsal bilgileri erişime açarak erişilebilirliği, yanıt verimliliğini ve kullanıcı memnuniyetini artırmıştır.

Pratik katkılar: Öğrenci işleri portalları ve mobil uygulamalara entegrasyon, idari iş yükünü azaltabilir.

Özgünlük: Çalışma, yükseköğretimde bağlama duyarlı dijital öğrenci hizmetleri için RAG kullanımını göstermektedir.

Anahtar Kelimeler: Yapay Zekâ, Büyük Dil Modelleri, Bilgi Getirimiyle Artırılmış Üretim, Doğal Dil İşleme, Dijital Öğrenci Hizmetleri

JEL Sınıflandırması: I23, O33

* Ankara Yıldırım Beyazıt Üniversitesi, m.f.ozer@aybu.edu.tr, 0000-0002-5769-0204

** Ankara Yıldırım Beyazıt Üniversitesi, mtakpinar@aybu.edu.tr, 0000-0003-4651-7788

*** Ankara Yıldırım Beyazıt Üniversitesi, sceyhan@aybu.edu.tr, 0000-0001-5721-6855

Araştırma Makalesi (Research Article)
Geliş (Submitted) : 29 Nisan 2026
Düzeltme (Revision) : 5 Haziran 2026
Kabul (Accepted) : 12 Haziran 2026

The study is supported by Ankara Yildirim Beyazit University BAP Departmant (Project No: 2764).

1. Introduction

The rapid advancement of digitalization and AI has transformed higher education, necessitating the development of intelligent and adaptive solutions to enhance student services. Traditional university administrative processes are often hindered by inefficiencies, delays, and limited accessibility, impeding students' ability to obtain timely and accurate information. AI-powered chatbots LLMs and NLP have emerged as a promising alternative, enabling automated, interactive, and scalable service delivery (Hadi et al., 2023). However, conventional rule-based and static chatbot systems are inherently constrained by predefined datasets, resulting in outdated or contextually inadequate responses that fail to address evolving student needs (Izadi & Forouzanfar, 2024). To overcome these limitations, this study proposes an AI-driven chatbot system based on the RAG model, which combines real-time knowledge retrieval with generative AI capabilities. Unlike static models that rely solely on pre-trained linguistic patterns, RAG dynamically accesses external databases, ensuring accurate, up-to-date, and contextually relevant responses to student inquiries. This approach significantly enhances chatbot adaptability, reducing reliance on static datasets while improving responsiveness and reliability (Jurafsky & Martin, 2024).

Various techniques in LLM optimization serve distinct purposes in AI-driven applications (Alto, 2024). Fine-tuning allows models to specialize in specific domains, whereas prompt engineering refines output quality without additional training. Few-shot and zero-shot learning facilitate adaptability in novel tasks with minimal labeled data, while reinforcement learning from human feedback (RLHF) enhances model reliability and ethical alignment (Sarkar, 2025). Model distillation reduces computational demands by transferring knowledge from complex architectures to lightweight models, and the Mixture of Experts (MoE) framework optimizes efficiency through specialized sub-models. Empirical studies underscore the efficacy of LLM-enhanced chatbots. Chen and colleagues (2024) reported a 94.7% accuracy rate for a GPT-3.5-based chatbot in university admissions, while Olawore and colleagues (2024) demonstrated the superior performance of re-trieval-augmented models over standalone LLMs. Despite these advancements, LLMs remain opaque "black box" systems, posing challenges in explainability, bias mitigation, and hallucination (misleading information generation). Address-ing these concerns requires robust evaluation frameworks to ensure AI-driven chatbots are transparent, ethical, and contextually precise (Zhao et al., 2024).

This study develops and evaluates a RAG-based chatbot for university student affairs, deployed in a pilot university in Türkiye. Through interaction logs and user satisfaction surveys, the research assesses its effectiveness in enhancing student engagement, accessibility, and administrative efficiency, contributing to the broader digital transformation of higher education services.

2. Literature Review

The evolution of AI has been driven by the development and assessment of increasingly sophisticated models. The evaluation of AI systems plays a crucial role in identifying their limitations and improving their performance, ultimately leading to the creation of more robust and reliable models (Sarker, 2022). In recent years, LLMs have garnered significant attention across academic and industrial domains due to their impressive capabilities in Natural Language Processing, machine translation, text summarization, and conversational AI. Their ability to process vast amounts of textual data and generate human-like responses has positioned them as a key technology in numerous applications, particularly in ed-ucation, where they support students in accessing critical information. (Kamath et al., 2024).

Moreover, models like GPT-4 are trained using RLHF, improving their ability to generate more ethically aligned and user-appropriate outputs. However, despite their remarkable capabilities, LLMs face critical challenges, including their reliance on static knowledge, the risk of hallucinations (generating false or misleading information), and their limited access to domain-specific data sources. These issues raise concerns regarding the accuracy, reliability, and ethical implications of deploying LLMs in real-world applications (hang et al., 2024).

To address these limitations, RAG has emerged as a promising approach that enhances LLM performance by integrating knowledge retrieval mechanisms (Gao et al., 2023). nlike traditional LLMs that rely solely on pre-trained knowledge, RAG-based models dynamically retrieve relevant and up-to-date information from external databases, academic documents, and structured knowledge sources. This approach significantly mitigates the risk of outdated or incorrect responses, as the chatbot can access real-time institutional data, ensuring more accurate and context-aware outputs. The RAG framework consists of two primary components:(1) the retrieval phase, where user queries are matched with relevant documents in a designated knowledge base, and (2) the generation

phase, where the retrieved information is processed alongside the LLM to generate precise and contextually enriched responses (Chang et al., 2024).

Table 1: Limitations of Large Language Models (LLMs) and the Potential of Retrieval-Augmented Generation (RAG) for Mitigation

Weakness of LLMs	Can it be mitigated with RAG?	Explanation	Sources
Lack of real-time knowledge	Yes	RAG retrieves up-to-date information from external sources such as the internet, databases, or private documents to generate responses.	Brown et al. (2020)
Hallucination (Generating Misleading Information)	Partially	RAG can reduce hallucinations if it retrieves data from reliable sources. However, if incorrect or misleading data is retrieved, LLMs may still generate false information.	Petroni et al. (2019)
Insufficient Context Understanding	Yes	RAG helps LLMs establish better contextual understanding by retrieving relevant information from long texts or specific sources.	Fan et al. (2024)
Non-updatable Static Knowledge Base	Yes	While LLMs remain static after training, RAG enables dynamic data retrieval, ensuring access to updated content.	Guu et al. (2020)
Weaknesses in Logical and Abstract Reasoning	Partially	RAG can improve decision-making processes by providing more data, but it does not directly enhance logical reasoning capabilities.	Zhao et al. (2024a)
Biases and Ethical Concerns	Not completely	If RAG retrieves data from biased sources, LLM- generated responses may still be problematic. Therefore, careful selection of sources is required.	Zhao et al. (2024b)

Table 1 outlines the key limitations of LLMs and assesses the extent to which the RAG model can mitigate these challenges. While RAG effectively addresses

real-time knowledge gaps and improves context understanding by dynamically retrieving up-to-date information, its ability to mitigate hallucinations and logical reasoning weaknesses remains partial, as it depends on the reliability of external sources. Additionally, RAG helps overcome the static nature of LLM knowledge bases, enabling access to dynamic content. However, biases and ethical concerns persist, as RAG cannot fully eliminate the risk of generating biased responses if the retrieved data itself is flawed. These findings highlight the strengths and limitations of RAG as a complementary approach to enhancing LLM performance in real-world applications.

The integration of RAG and LLMs is particularly advantageous for chatbot applications that require high accuracy and dynamic information access. Traditional chatbot systems, often rule- based or dependent on static knowledge repositories, are unable to adapt to evolving contexts, making them unsuitable for domains requiring real-time updates, such as university student affairs. In contrast, RAG-based chatbots leverage both generative language capabilities and real-time knowledge retrieval, offering a more reliable, informative, and user-friendly experience (hang et al., 2024). This study proposes the development of a RAG-enhanced chatbot for university student services, designed to provide students with quick and accurate access to university regulations, academic pol-icies, and administrative support. The chatbot combines the linguistic fluency of LLMs with RAG's retrieval mechanisms, ensuring that responses are not only grammatically coherent but also factually precise. The evaluation of this chatbot focused on its accuracy, user satisfaction, and effectiveness in facilitating access to student services, thereby contributing to the broader digital transformation of higher education administration.

3. Methodology

This study focuses on the development and evaluation of an AI-driven chatbot system for university student affairs, employing the RAG model to enhance response accuracy and accessibility. Unlike conventional chatbots that rely on static datasets, this system integrates real-time knowledge retrieval mechanisms with LLMs, ensuring that responses remain up to date and contextually relevant (Jeon et al., 2025). By dynamically sourcing information from institutional databases, the chatbot aims to facilitate seamless student interactions concerning administrative services, academic procedures, and general student affairs while

improving the overall efficiency of university support systems. Methodologically, the study was designed as an applied system-development and evaluation study, combining qualitative needs analysis with post-implementation user evaluation. This design enabled the research to both identify student-service problems before system development and assess the usability and performance of the chatbot after deployment.

To align the chatbot's design with students' actual needs, an initial qualitative analysis was conducted through preliminary interviews with 20 students from various faculties. These interviews provided critical insights into the most pressing issues students face when engaging with administrative services, including delays in response times, difficulties in accessing accurate information, and the absence of a personalized guidance mechanism. The data gathered from these interviews informed the chatbot's functional and structural design, ensuring it effectively addresses students' primary concerns. The preliminary interviews were used as the needs-analysis stage of the study. Student responses were examined to identify common problems and frequently requested information categories, and these findings were used to guide the chatbot's content and functional design.

A diverse participant pool was used to evaluate the chatbot's functionality. Students were randomly selected from different faculties to assess the chatbot's effectiveness across various academic disciplines. Participants interacted with the chatbot in real-world scenarios, using its features to obtain academic and administrative information. Their interactions were systematically recorded and analyzed based on key performance metrics, including response accuracy, processing speed, contextual coherence, and overall usability. This structured evaluation provided a comprehensive assessment of the chatbot's performance across different student demographics and information needs. In addition to students, relevant academic and administrative personnel also tested the chatbot by asking unit-specific questions related to their own faculties, departments, and service areas. This allowed the evaluation to cover both student-centered scenarios and institutionally differentiated information needs.

The chatbot architecture consists of multiple interdependent components, each designed to optimize interaction quality and information accessibility. At its core, the chatbot is integrated with a structured university database, compris-

ing official institutional resources such as academic regulations, frequently asked questions (FAQ), academic calendars, course registration guidelines, scholarship and housing policies, and other essential student services. Upon receiving a query, the system retrieves the most relevant documents from this repository, leveraging RAG's retrieval mechanism to provide real-time, contextually appropriate responses (Ramalingam, 2023).

To further refine response quality, the retrieved data is processed through an LLM, which enhances the chatbot's linguistic fluency and contextual adaptability. This hybrid model ensures that the chatbot not only delivers accurate factual information but also generates coherent, human-like responses that improve user engagement. The chatbot was deployed via both the university's student affairs web portal and a mobile application, maximizing accessibility and usability for students.

To measure the chatbot's effectiveness, interaction logs were analyzed by focusing on response precision, efficiency, and student satisfaction. Participants were encouraged to provide immediate feedback after each interaction, facilitating real-time system refinements. Additionally, a structured questionnaire was administered to assess key usability metrics, including response accuracy, relevance, processing time, and overall user experience. The quantitative evaluation data were analyzed descriptively through usability scores, satisfaction percentages, and average response time values. The qualitative feedback obtained from interviews and user comments was examined to identify recurring perceptions regarding accessibility, clarity, response usefulness, and the overall user experience.

In the final evaluation phase, a second round of interviews was conducted with 40 university students to further assess the chatbot's variety, reliability, and consistency in delivering accurate responses. The participants consisted of 23 female and 17 male university students aged between 18 and 25, representing various academic departments, including Management Information Systems, Business Administration, Political Science and Public Administration, Medicine, and other related fields. This phase provided deeper insights into how well the chatbot adapted to diverse student inquiries and whether any refinements were needed for future iterations. By integrating both preliminary and post-development qualitative evaluations, this study established a comprehensive framework for AI-enhanced digital student services and contributed to the broader digital transformation of university administrative systems.

3.1. Ethical Approval

This study was approved by the Social and Humanities Sciences Ethics Committee of Ankara Yıldırım Beyazıt University (Research Serial Number: 793; Decision Date and Number: 22.09.2025 – 07/793). Informed voluntary consent was obtained from all participants, and the study was conducted in accordance with established ethical guidelines for social science research.

4. Findings and Discussion

The preliminary interviews conducted with twenty students before system development identified the main problems related to existing university service platforms and provided valuable insights into students' access to information and satisfaction with administrative services. The initial phase of this study revealed that students predominantly rely on the university's official website for information. However, despite being the most frequently used resource, the website is not always user-friendly, making it difficult for students to locate the necessary information efficiently. While the official website is considered a reliable source, issues such as poor content organization and overly formal language may hinder students' ability to quickly comprehend the information they seek.

Table 2: Students' Primary Channels for Accessing University Information

Information Source	Percentage of Students (%)
University's official website	80%
Faculty or academic advisors	45%
Social media platforms	35%
Student affairs offices	20%

In addition to the university's website, students frequently turn to their professors and academic advisors for guidance on academic matters. This approach allows for personalized and tailored responses; however, it may not be equally accessible to all students, particularly those who are introverted or find it difficult to engage in direct communication. In terms of administrative processes, consulting the student affairs office is a commonly adopted method, yet the findings

indicate that accessing student affairs services presents significant challenges. A considerable number of respondents reported difficulties in reaching student affairs personnel via phone or email, with extended response times being a common concern. Furthermore, in-person visits often involve long waiting times and overcrowding, which further complicate students' ability to obtain information efficiently.

Given the challenges associated with accessing information and the existing limitations of current systems, it becomes evident that universities need to implement more effective solutions for streamlining information-sharing processes. At this point, the development of an AI- powered chatbot system emerges as a highly viable solution to address the fundamental issues students encounter in accessing academic and administrative information. Considering the challenges associated with navigating university websites, contacting student affairs offices, and seeking guidance from academic advisors, an NLP-based chatbot could provide students with instant, 24/7 responses while ensuring that the information provided is accurate and up to date. Such a system could alleviate the workload of student affairs offices by handling frequently asked questions automatically, while also offering a dynamic and adaptive mechanism for addressing individualized queries.

This interpretation is consistent with previous systematic reviews showing that chatbots in educational settings can provide rapid, personalized, and accessible support for students and institutional staff. In particular, educational chatbot studies indicate that such systems are frequently used to answer routine questions, support student guidance, reduce uncertainty in communication processes, and decrease repetitive administrative workload (Labadze et al., 2023).

Moreover, a RAG-based chatbot model was not restricted to pre-programmed responses but instead retrieved information from up-to-date university databases and official sources, thereby ensuring greater accuracy and relevance. This enabled students to efficiently access crucial information related to course registration, class schedules, financial aid, scholarships, and academic calendars without unnecessary delays. Additionally, such a system contributed to improving operational efficiency by reducing the volume of individual inquiries directed at student affairs personnel, allowing human resources to focus on more complex administrative tasks.

Table 3*:* Post-Test Evaluation Metrics of the RAG-Based Chatbot

Evaluation Metric	Result
Usability	Mean = 84.2
Satisfaction	87%
Average Response Time	3.2 seconds

Usability scored a mean of 84.2 based on the post-implementation usability evaluation, while satisfaction reached 87%. Average response time was approximately 3.2 seconds per query. As presented in Table 3, these results demonstrate that the RAG-based chatbot achieved high usability, strong user satisfaction, and efficient performance.

In line with the task-based nature of usability testing, the post-implementation evaluation was structured around scenario-based questions reflecting common academic and administrative information needs. These scenarios were not limited to general student satisfaction; rather, students were directed to use the chatbot for specific information-seeking situations, while relevant administrative and academic personnel also posed unit-specific questions related to their own service areas. In this way, the evaluation covered different institutional contexts, including faculty-level issues, department-specific inquiries, and administrative procedures handled by different student-service units. This approach enabled the analysis to focus on whether the chatbot could provide accurate, clear, and timely responses across different types of real university-service scenarios.

The scenario-based evaluation included different categories of academic and administrative information needs. In the academic calendar and deadline-related scenarios, students asked questions about registration periods, add-drop processes, examination dates, and other time-sensitive procedures. These questions made it possible to evaluate whether the chatbot could retrieve accurate and up-to-date institutional information and present it in a clear and timely manner. Similarly, course registration scenarios focused on students' ability to obtain procedural guidance about course selection, registration problems, and related faculty or department-level requirements. In these cases, the participants generally indicated that the chatbot provided more direct and understandable guidance than fragmented website pages.

The evaluation also included scenarios related to scholarships, financial support, document requests, petitions, applications, and other student affairs procedures. These questions were important because they reflected information needs that often require students to contact different administrative units. Feedback from these scenarios suggested that the chatbot helped reduce uncertainty by presenting official requirements and procedural steps in a more accessible form. In addition to student-generated questions, relevant academic and administrative personnel posed unit-specific questions concerning their own faculties, departments, and service areas. This enabled the evaluation to consider whether the chatbot could respond consistently to institutionally differentiated inquiries rather than only answering general student questions.

Furthermore, post-implementation interviews with students revealed that the chatbot's simple interface and rapid interaction capabilities made it highly user-friendly. The majority of participants emphasized that the chatbot's guidance was clearer and more comprehensible than that provided by the university website. Students also highlighted that the chatbot offered a more accessible and less intimidating alternative to direct communication with faculty or administrative staff, which was particularly beneficial for individuals reluctant to engage in face-to-face interactions. In addition, the system's ability to provide immediate and accurate responses substantially reduced dependence on delayed email correspondence and overcrowded student affairs offices. Compared to existing channels, the chatbot was perceived as more efficient, equitable, and user-centered, enabling students to access essential academic and administrative information promptly while eliminating barriers associated with traditional processes.

Moreover, prior to the chatbot's deployment, students frequently encountered inconsistencies and outdated information across different administrative channels, which undermined confidence in the reliability of institutional communication. The process of contacting student affairs was also reported as slow and inconvenient, often involving long waiting times and delayed responses via email or phone.

These issues not only limited accessibility but also negatively shaped perceptions of service efficiency. Traditional channels were further characterized as frustrating, time-consuming, and lacking user-friendliness, which contributed to diminished overall satisfaction. Following the integration of the RAG-based chat-

bot, these challenges were largely resolved, as the system consistently provided accurate, institutionally verified information, ensured instant access with significantly shorter response times, and delivered a more seamless and user-friendly experience. Consequently, students reported higher levels of satisfaction and demonstrated a clear willingness to rely on the chatbot for future academic and administrative inquiries.

4.1. Limitations of the Study

Although the findings demonstrate the potential of a RAG-based chatbot for improving access to digital student services, several limitations should be acknowledged. First, the study was conducted in a pilot university context in Türkiye; therefore, the findings may not be directly generalizable to all higher education institutions with different administrative structures, student profiles, or digital infrastructures. Second, the evaluation was based on a relatively limited participant group, consisting of preliminary interviews with 20 students and post-implementation evaluation with 40 students. While this sample provided valuable insights into student needs and usability perceptions, larger and more diverse samples would strengthen the external validity of future studies.

Another limitation concerns the scope of the usability evaluation. Although the assessment included scenario-based questions from students as well as unit-specific questions from relevant academic and administrative personnel, these scenarios mainly focused on frequently encountered academic and administrative information needs, such as course registration, academic calendars, scholarship information, document requests, and student affairs procedures. (Huang et al., 2025; Zhang & Zhang, 2025). Therefore, the chatbot's performance in highly complex, personalized, legally sensitive, or confidential cases was not fully examined. In addition, the study primarily relied on user feedback, interaction logs, response time, and perceived satisfaction; future research could include more detailed experimental comparisons with existing service channels and longitudinal measures of actual administrative workload reduction. Finally, as with all RAG-based systems, response quality depends on the accuracy, completeness, and timely updating of the institutional knowledge base. If the underlying documents are outdated, incomplete, or inconsistent, the chatbot may still generate responses that require verification by authorized university personnel.

5. Conclusion

In conclusion, the integration of AI-driven chatbot solutions should be regarded not merely as a technological enhancement but as a strategic necessity for universities aiming to optimize information-sharing systems, strengthen student satisfaction, and ensure a more accessible and efficient communication framework. Future research will focus on assessing the long-term effectiveness of chatbot-based models in enhancing students' access to academic and administrative information and on identifying the most effective implementation strategies for higher education institutions.

References

Alto, V. (2024). *Building LLM powered applications: Create intelligent apps and agents with large language models*. Packt Publishing.

Antico, M., Lopez, P., & Rivera, C. (2024). Enhancing student support with RAG-powered chatbots: A case study in higher education. *Education and AI Journal, 22*(4), 77–92.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877–1901.

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Yu, P. S., Yang, Q., & Xie, X. (2024). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology, 15*(3), Article 39.

Chen, Z., Zou, D., Xie, H., Lou, H., & Pang, Z. (2024). Facilitating university admission using a chatbot based on large language models with retrieval-augmented generation. *Educational Technology & Society, 27*(4), 454–470.

Fan, W., Ding, Y., Ning, L., Wang, S., Li, H., Yin, D., ... & Li, Q. (2024, August). A survey on RAG meeting LLMs: Towards retrieval-augmented large language models. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 6491–6501).

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., ... & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. *arXiv preprint* arXiv:2312.10997.

Guu, K., Lee, K., Tung, Z., Pasupat, P., & Chang, M. (2020). REALM: Retrieval-augmented language model pre-training. In *Proceedings of the 37th International Conference on Machine Learning* (pp. 3930–3942).

Hadi, M. U., Qureshi, R., Shah, A., Irfan, M., Zafar, A., Shaikh, M. B., ... & Mirjalili, S. (2023). Large language models: A comprehensive survey of its applications, challenges, limitations, and future prospects. *Authorea Preprints*, 1–26.

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., ... & Liu, T. (2025). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, *43*(2), 1-55.

Izacard, G., & Grave, E. (2021). Leveraging passage retrieval with generative models for open-domain question answering. *arXiv preprint*.

Izadi, S., & Forouzanfar, M. (2024). Error correction and adaptation in conversational AI: A review of techniques and applications in chatbots. *AI, 5*(2), 803–841.

Jeon, J., Sim, Y., Lee, H., Han, C., Yun, D., Kim, E., ... & Lee, J. (2025). ChatCNC: Conversational machine monitoring via large language model and real-time data retrieval augmented generation. *Journal of Manufacturing Systems, 79*, 504–514.

Jurafsky, D., & Martin, J. H. (2024). *Speech and language processing* (3rd ed., draft). Stanford University & University of Colorado at Boulder.

Kamath, U., Keenan, K., Somers, G., & Sorenson, S. (2024). *Large language models: A deep dive*. Springer.

Karpukhin, V., Oguz, B., Min, S., Lewis, P., Wu, L., Edunov, S., ... & Yih, W. (2020). Dense passage retrieval for open-domain question answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing* (pp. 6769–6781).

Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: systematic literature review. *International journal of Educational Technology in Higher education*, *20*(1), 56.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Riedel, S. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems, 33*, 9459–9474.

Logeswaran, L., Lee, H., Devlin, J., Lee, K., Toutanova, K., & Gardner, M. (2019). Zero-shot entity linking by reading entity descriptions. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 3449–3460).

Neupane, R., & Zhang, L. (2024). BARKPLUG V.2: A university chatbot for enhancing student services. *Mississippi State University AI Research Papers, 15*(1), 25–40.

Nguyen, T. P., & Quan, H. (2025). Unified hybrid RAG: An approach to improving university chatbot systems. *Artificial Intelligence in Higher Education, 18*(3), 50–68.

Olawore, K., McTear, M., & Bi, Y. (2024). Development and evaluation of a university chatbot using deep learning: A RAG-based approach.

Petroni, F., Rocktäschel, T., Lewis, P., Bakhtin, A., Wu, Y., Miller, A., & Riedel, S. (2019). Language models as knowledge bases? In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing* (pp. 2463–2473).

Ramalingam, S. (2023). *RAG in action: Building the future of AI-driven applications*. Libertatem Media.

Ranoliya, B. R., Raghuwanshi, N., & Singh, S. (2017). Chatbot for university-related FAQs. In *Proceedings of the International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 1525–1530).

Sarkar, U. E. (2025). Evaluating alignment in large language models: A review of methodologies. *AI and Ethics*, 1–8.

Sarker, I. H. (2022). AI-based modeling: Techniques, applications and research issues towards automation, intelligent and smart systems. *SN Computer Science, 3*(2), 158.

Shuster, K., Humeau, S., Bordes, A., & Weston, J. (2021). Retrieval augmented generation for knowledge-intensive NLP tasks. *arXiv preprint*.

Weidinger, L., Uesato, J., Mellor, J., Higgins, I., Chang, J., Huang, P.-S., ... & Gabriel, I. (2022). Ethical and social risks of harm from language models. *arXiv preprint*.

Weizenbaum, J. (1966). ELIZA: A computer program for the study of natural language communication between man and machine. *Communications of the ACM, 9*(1), 36–45.

Zhang, W., & Zhang, J. (2025). Hallucination mitigation for retrieval-augmented large language models: a review. *Mathematics*, *13*(5), 856.

Zhao, H., Chen, H., Yang, F., Liu, N., Deng, H., Cai, H., ... & Du, M. (2024). Explainability for large language models: A survey. *ACM Transactions on Intelligent Systems and Technology, 15*(2), 1–38.

Zhao, S., Yang, Y., Wang, Z., He, Z., Qiu, L. K., & Qiu, L. (2024). Retrieval-augmented generation (RAG) and beyond: A comprehensive survey on how to make your LLMs use external data more wisely. *arXiv preprint* arXiv:2409.14924.

Zhou, Y., Liu, Y., Li, X., Jin, J., Qian, H., Liu, Z., ... & Yu, P. S. (2024). Trustworthiness in retrieval-augmented generation systems: A survey. *arXiv preprint* arXiv:2409.10102.