ISSN 2409-2665
Journal of Logistics, Informatics and Service Science
Vol. 12 (2025) No. 7, pp. 56-67
DOI:10.33168/JLISS.2025.0704

Generative AI in University Customer Service: A Comprehensive Framework for Enhancing Efficiency and Experience

Aina Eirena, Nathar Shah

Multimedia University, Cyberjaya, Malaysia

nathar.packier@mmu.edu.my

Abstract. This research investigates the application of Generative Artificial Intelligence (GAI) in university customer service, specifically to automate the real-time answering of student questions. The study assesses the effectiveness of GAI solutions, including large language models (LLMs) and hybrid systems, for enhancing the responsiveness, scalability, and personalization of Student Support Systems (SSSs), while also examining the ethical and operational limits of these technologies. Through a systematic literature review, the study synthesizes key trends and identifies a gap in domain-specific, theoretically-grounded studies. An exploratory comparative experiment is then presented using a manually constructed dataset of 200 student inquiries to evaluate six different models across quantitative metrics (accuracy, F1-score, latency) and qualitative dimensions (response quality, empathy). The findings validate the potential of GAI, with GPT-4 significantly surpassing traditional and deep learning models in accuracy, F1-score, and user-perceived empathy. However, the study also highlights critical challenges, including hallucinations, bias, and privacy concerns, which necessitate transparent, secure, and inclusive application. The findings, interpreted through the lens of Service Quality Theory, validate GAI's potential to revolutionize university service provision while also drawing attention to the requirements for human oversight and continuous refinement.

Keywords: Generative AI, Customer Service, Large Language Models (LLMs), NLP, Service Quality.

1. Introduction

During the digital era, universities are required to adapt to the changing aspirations of a tech-smitten student body. Artificial Intelligence (AI) conversational agents (CA) or chatbots have emerged as technologies capable of providing automated customer service across various domains (Nicolescu & Tudorache, 2022; Wang, 2018; Khennouche et al., 2024; Misischia et al., 2022; Alkishri et al., 2025b). Generative AI (GAI), which leverages deep learning architectures to create novel and coherent data formats such as text, represents a paradigm shift in machine learning (Hagos et al., 2024; Gupta et al., 2024; Tan et al., 2024; Goodfellow et al., 2020). Fuelled by advancements in deep learning and computational resources, generative models push the boundaries of what machines can create (Tan et al., 2024; Hagos et al., 2024).

Large language models (LLMs) are a specific application of GAI that have shown remarkable performance on many Natural Language Processing (NLP) tasks (Hagos et al., 2024; Vatsal & Dubey, n.d.). LLMs play a foundational role in the generative capabilities of AI, demonstrating remarkable abilities in understanding and generating human language, opening up opportunities across a wide range of domains such as question answering and text summarization (Hagos et al., 2024; Tan et al., 2024). Prompt engineering plays a key role in leveraging the existing abilities of LLMs to achieve performance gains on various NLP tasks (Vatsal & Dubey, n.d.; Sahoo et al., 2024; Khennouche et al., 2024; Odede & Frommholz, 2024). Prompt engineering, defined as the art or technique of using prompts and instructions to tell LLMs what to do, allows users to intelligently extract LLMs' knowledge through a basic natural language conversational exchange (Vatsal & Dubey, n.d.; White et al., 2023; Reinhard et al., 2024). The potential applications of GAI in content creation are vast and transformative (Tan et al., 2024; Gupta et al., 2024).

In the context of customer service, GAI affords companies new possibilities to communicate, connect, and engage customers (Jensen et al., 2024; Reinhard et al., 2024). The efficiency and resource-saving benefits of AI-powered customer interaction are widely discussed (Khennouche et al., 2024; Nicolescu & Tudorache, 2022; Wang, 2018; Jensen et al., 2024; Misischia et al., 2022). However, integrating these advanced systems, such as advanced generative models like GPT-4, into the academic sector requires a strategic approach, particularly concerning factual accuracy and ethical deployment (OpenAI, 2023; Khennouche et al., 2024; Paudel & Acharya, 2024; Dwivedi et al., 2023).

This research aims to bridge the gap between the theoretical capabilities of Generative AI and its practical implementation in the unique context of university customer service. This includes:

• Investigating domain-specific prompt design tailored for academic customer service.

• Evaluating performance across multiple dimensions, including accuracy, response relevance, ethical compliance, and scalability.

By addressing these elements and applying a theoretical framework, this research seeks to enhance the understanding of how Generative AI can be effectively and responsibly implemented in the higher education sector.

2. Literature Review and Theoretical Framework

The integration of artificial intelligence into customer service has advanced rapidly across various sectors, including education. Higher education institutions are increasingly adopting AI solutions to enhance operational efficiency and improve student satisfaction. This section reviews the evolution of customer service technologies in academia and examines the core techniques used in sentiment analysis and inquiry classification, all while positioning our analysis within a robust theoretical framework. The formal literature review process employed a systematic literature review (SLR) approach (Nicolescu & Tudorache, 2022; Kitchenham, 2004; Okoli, 2015; Webster & Watson, 2002; vom Brocke et al., 2015).

2.1. The Evolution of AI in Higher Education Customer Service

Historically, university customer service was primarily managed through manual methods such as in-person consultations, email, and static FAQs. These methods were resource-intensive and often resulted in delayed or mismatched responses, particularly during peak times. The initial move toward automation involved the implementation of rule-based chatbots and ticketing platforms, which improved response times but were still limited in their ability to handle nuanced inquiries. Conversational agents or chatbots have been defined in different ways (Nicolescu & Tudorache, 2022). A chatbot is a computer program that interacts with people (Ahmad et al., 2021; Shawar & Atwell, 2007). It uses computer programs and algorithms to perform semantic analysis and provide appropriate responses, conducting a conversation in a chat application (Ahmad et al., 2021). Early conversational systems relied on rule-based pattern matching, such as ELIZA and SCHOLAR (Gupta et al., 2024; Carbonell, 1970). Chatbots are increasingly finding their way into e-commerce and e-services, offering opportunities to improve customer service (Misischia et al., 2022).

The next phase introduced AI-driven chatbots based on Natural Language Processing (NLP), but early versions relied on deterministic algorithms that restricted their conversational depth. The emergence of transformer-based models enabled a more dynamic and flexible interaction, paving the way for advanced virtual assistants in educational settings. Generative AI (GAI) focuses on synthesizing content that is often indistinguishable from human-generated content (Gupta et al., 2024). Fuelled by advancements in deep learning (LeCun et al., 2015; Hagos et al., 2024), generative models push the boundaries of what machines can create (Tan et al., 2024). LLMs are a specific application of GAI that show remarkable performance on NLP tasks, demonstrating abilities in understanding and generating human language (Hagos et al., 2024). Advanced LLMs like GPT-4 possess capabilities that surpass previous models (OpenAI, 2023).

In the academic domain, AI CAs are utilized for various purposes, including admissions (Ahmad et al., 2021), academic advising (Ranoliya et al., 2017; Alkishri et al., 2025a), and addressing FAQs (Khennouche et al., 2024). These systems aim to reduce workload for human staff, save time, and improve the student experience (Alkishri et al., 2025a; Odede & Frommholz, 2024). For example, studies have shown the potential of AI in university admissions (Ahmad et al., 2021; Chen et al., 2019). However, AI implementation in this sector requires rigor due to the importance of factual accuracy (OpenAI, 2023).

2.2. Theoretical Foundations: A Framework for Evaluation

To move beyond a purely technical comparison of model performance, this research is grounded in the established principles of service science. Our analysis is informed by the **Service Quality (SERVQUAL) Theory**, which provides a foundational framework for evaluating service delivery. SERVQUAL posits that customer perception of service quality is measured across several key dimensions, three of which are directly relevant to our study:

Responsiveness, **Assurance**, and **Empathy**.

- **Responsiveness:** This dimension measures the willingness to provide prompt and helpful service. It aligns with our quantitative metrics of latency and the overall efficiency of the AI solution.
- **Assurance:** This relates to the perceived knowledge and trustworthiness of the service provider. In our context, this corresponds to the accuracy, factual correctness, and clarity of the AI-generated responses.
- **Empathy:** This is the perception of caring, individualized attention. Our qualitative assessment of the AI's tone and its ability to provide supportive language is a direct measure of this crucial dimension.

2.3. Generative AI in Customer Service: Capabilities and Challenges

Generative AI, particularly LLMs like GPT, has revolutionized customer interaction by offering dynamic, human-like responses. Unlike retrieval- or rule-based systems, GAI can generate new information, rewrite responses, and adapt to conversation styles, which is highly beneficial for handling diverse topics like course registration and financial aid with minimal pre-programming. The ability to customize a model to adhere to institutional language and policies through timely engineering and tweaking is a key advantage.

However, the transformative potential of GAI is tempered by several significant challenges. These models can generate inaccurate or fabricated information (hallucinations), struggle with domain-specific accuracy, and raise ethical concerns related to data bias, privacy, and accountability. These issues underscore the need for hybrid solutions that combine GAI with traditional retrieval mechanisms and rule-based filters.

2.4. The Research Gap: A Thematic Synthesis of Existing Work

Previous comparative studies have largely focused on other domains, as summarized in Table 1. ChatGPT outperformed traditional methods in the finance sector (Misischia et al., 2022; Ji & Zhang, 2022; Li et al., 2022; Guia et al., 2019; Ranoliya et al., 2017), while Naive Bayes and SVM were viable but unable to handle subtle emotional expressions (Guia et al., 2019; Ranoliya et al., 2017; Ahmad et al., 2021). GPT-based models in customer support and stressed the significance of context-aware generation and ethical choices (Khennouche et al., 2024; Paudel & Acharya, 2024; Dwivedi et al., 2023).

Table 1: Systematic Literature Review Synthesis

Study/Reference	AI Technique(s)	Application	Contribution/Limitation Highlighted
Ahmad et al. (2021)	Pre-trained Language Model, Response Ranking Model	University Admissions	Shows potential for AI in admissions, but lacks a broader comparative analysis across models (Ahmad et al., 2021; Chen et al., 2019).
Ranoliya et al. (2017)	Chatbots (AIML, LSA)	Academic Services	Highlights the evolution and limitations of rule-based systems in an academic context (Ranoliya et al., 2017; Alkishri et al., 2025b; Alkishri et al., 2025a).
Wang (2018)	Intelligent Systems (NLP/ML/BD)	E-Commerce	Demonstrates impact on user experience but does not address the unique needs of academia (Wang, 2018; Li et al., 2022).

A notable gap remains in studies that specifically target customer service in educational settings. Most research treats sentiment analysis and response generation as separate tasks, whereas university applications require an integrated approach that can handle academic jargon, policy interpretation, and multilingual support. This study aims to fill that void by:

- Conducting a comparative evaluation of GAI models with traditional machine learning and lexicon-based techniques for student inquiries.
- Investigating domain-specific prompt design tailored for academic customer service.
- Evaluating performance across multiple dimensions, including accuracy, response relevance, ethical compliance, and scalability.

By addressing these elements and applying a theoretical framework, this research seeks to enhance the understanding of how Generative AI can be effectively and responsibly implemented in the higher education sector.

3. Methodology

This section introduces the detailed process of our research, which consists of a systematic literature review (SLR) and a controlled experiment. The methodology is designed to bridge the theoretical background with empirical validation, ensuring that the experiment reflects real-world academic service scenarios.

3.1. Research Design

The study employs a mixed-method approach. A Systematic Literature Review (SLR) was conducted to analyze trends, classify AI techniques, and identify research gaps in the domain of AI-based university customer service (Nicolescu & Tudorache, 2022; Kitchenham, 2004; Okoli, 2015; Webster & Watson, 2002; vom Brocke et al., 2015). SLR is characterized as a way to identify, evaluate, and interpret all available research relevant to a research topic (Nicolescu & Tudorache, 2022; Kitchenham, 2004). The analysis aimed at reviewing existing perceptions (Reinhard et al., 2024) and involved a full assessment of relevant literature. A controlled experiment was then designed to test the effectiveness of selected models—especially Generative AI—in automating student inquiry responses.

3.2. Systematic Literature Review

The SLR was conducted to map the academic landscape by gathering relevant academic articles, journals, and conference proceedings from trusted digital libraries including IEEE Xplore, ACM Digital Library, ScienceDirect, and Google Scholar (Khennouche et al., 2024; Nicolescu & Tudorache, 2022). The rigorous review process followed established protocols (Khennouche et al., 2024; Nicolescu & Tudorache, 2022; Kitchenham, 2004; Okoli, 2015; Page et al., 2021). The search terms used were: "Generative AI in education," "university chatbot models," "sentiment analysis in student queries," "LLM-based customer service," and "automated helpdesk systems". Studies published between 2018 and 2024 were prioritized to ensure technical relevance and alignment with current AI trends.

• **Inclusion and Exclusion Criteria**: Specific criteria were established to ensure the relevance and quality of the literature (Nicolescu & Tudorache, 2022; Okoli, 2015). We selected peer-reviewed articles and conference papers that were related to customer service automation, with a specific focus on educational environments (Nicolescu & Tudorache, 2022). We also included works that explicitly mentioned generative AI, sentiment analysis, or intent classification and provided empirical results with measurable evaluation metrics (Nicolescu & Tudorache, 2022). To maintain a consistent interpretation, only papers published in English were included (Nicolescu & Tudorache, 2022). Unpublished sources, papers not available in full text, and works not generalizable to the specific problems of university customer service were excluded (Nicolescu & Tudorache, 2022; Ranoliya et al., 2017; Ahmad et al., 2021; Wang, 2018).

3.3. Experiment Design

The experimental component of the study was designed to assess the viability of GAI for university services using a diverse range of models and inquiries from real academic settings (Ahmad et al., 2021; Ranoliya et al., 2017; Wang, 2018). The inclusion of a broad range of models, from lexicon-based to state-of-the-art generative AI, was essential to provide a comprehensive and robust comparative analysis of different AI paradigms in this specific domain.

3.4. Dataset Construction and Validation

A custom dataset was constructed using Frequently Asked Questions (FAQs) from official university websites and a series of manually crafted queries. These handcrafted queries were designed to cover common academic categories such as admissions, course registration, graduation, fees, and technical support. The queries were also formulated to incorporate a range of unstructured complexities and sentiment tones, including those reflecting frustration or confusion, to better simulate real-world student interactions. This dataset, though limited in size to 200 queries, served as a foundational benchmark for our comparative evaluation.

3.5. Models Compared

This subsection details the selection of different conversational models used in the experiment, spanning traditional symbolic methods (e.g., lexicon-based approaches (Ji & Zhang, 2022; Guia et al., 2019)) to contemporary generative models (e.g., GPT-4 (OpenAI, 2023)). The following models were evaluated: a Lexicon-based model (VADER), two Machine Learning models (Naive Bayes and SVM), a Deep Learning model (Bi-LSTM), a Transformer model (BERT fine-tuned), and two Generative AI models (GPT-3.5 and GPT-4). Each model was tasked with interpreting student queries, identifying sentiment and intent, and generating a response where applicable.

3.6. Prompt Engineering (for GAI)

For the generative models (GPT-3.5 and GPT-4), prompt engineering was a critical component of the methodology. The prompts were carefully designed using few-shot examples and instruction-tuned queries to ensure that responses reflected the formal and supportive tone expected in a university helpdesk environment. This systematic process was essential for optimizing the models' ability to handle emotionally complex and ambiguous inputs while aligning their output with institutional standards.

3.7. Evaluation Metrics

The criteria used to evaluate the performance of the various models were rigorously defined, including quantitative metrics such as accuracy, fluency, and coherence, as well as subjective metrics such as response relevance and user experience (Følstad & Brandtzaeg, 2020; Følstad & Taylor, 2021). Performance was assessed using a combination of quantitative and qualitative metrics. Quantitative metrics included **Accuracy** (correctness of intent and sentiment classification), **F1-Score** (a balance of precision and recall), and **Latency** (response time). Qualitative metrics included **Response Quality** (assessed by human evaluators for relevance, clarity, and tone) and **Ethical Compliance** (screening for hallucinations and inappropriate language). Ethical compliance was measured through manual review of a subset of outputs to verify factual correctness and absence of harmful or biased content. Ethical compliance (screening for hallucinations and inappropriate language) was measured through manual review of a subset of outputs to verify factual correctness and absence of harmful or biased content (OpenAI, 2023). This step is crucial for ensuring the responsible deployment of the system (Dwivedi et al., 2023; Khennouche et al., 2024).

4. Results and Discussion

The systematic literature review of 30 selected papers identified several key trends. A notable trend is the increasing adoption of Generative AI, with universities transitioning from simple rule-based chatbots to more advanced models like GPT, T5, and PaLM to enhance student support systems (Hagos et al., 2024; Gupta et al., 2024; Tan et al., 2024). The review also confirmed that LLMs have shown remarkable performance on many different NLP tasks (Hagos et al., 2024; Vatsal & Dubey, n.d.). The review also revealed a persistent gap in domain-specific training for many GAI models, as they are often trained on general datasets that limit their ability to handle university-specific queries.

Furthermore, most studies focus on FAQs, with fewer addressing multilingualism, inclusivity, or local institutional needs.

The descriptive analysis covered 30 studies that empirically researched customers' experiences with customer service conversational agents (Nicolescu & Tudorache, 2022). This involved examining characteristics such as the year of publication, countries of origin of authors, countries where the empirical research was conducted, subject area of publication venues, research methods adopted, and industries involved in the studies (Nicolescu & Tudorache, 2022). For instance, core publication venues included journals from information systems and computing-related domains (over 55%), with Europe representing a pole of research on AI CA/chatbots in customer service (Nicolescu & Tudorache, 2022).

The narrative description focused on answering the study's research questions using findings aggregated from the analyzed publications (Nicolescu & Tudorache, 2022). It utilized a theoretical framework with seven constructs, detailing related publications, main findings, and implications for each construct (Nicolescu & Tudorache, 2022).

Research indicates that the overall customer experience with AI CAs is influenced by various factors (Nicolescu & Tudorache, 2022). These factors are typically grouped into three major categories: CA-related, user-related, and context-related factors (Nicolescu & Tudorache, 2022). Some studies focused on a single factor, such as CA/chatbot social presence, personality, or problem resolution capacity, while others considered a combination of factors (Nicolescu & Tudorache, 2022).

The application of conversational AI (CA) in education shows that chatbots are increasingly utilized for formative assessment, virtual tutoring, and administrative support, particularly within higher education settings (Alkishri et al., 2025b; Nicolescu & Tudorache, 2022; Carbonell, 1970; Khennouche et al., 2024; Alkishri et al., 2025a). Furthermore, understanding the factors that influence students' behavioral intentions to use generative AI tools like ChatGPT, including the mediating role of attitude, is crucial for improving their application in academic contexts (Paudel & Acharya, 2024).

4.1. Experimental Results

A comparative evaluation of the six different models was conducted using the 200-query benchmark dataset. The performance results are summarized in Table 2.

Table 2: Key Metrics Result

Model	Accuracy (%)	F1-Score	Response Quality (1–5)	Avg. Latency (s)
VADER	66.4	0.58	—	0.2
Naive Bayes	74.2	0.70	—	0.3
Bi-LSTM	83.7	0.82	—	1.5
BERT (Fine-tuned)	87.5	0.86	—	1.8
GPT-3.5	90.1	0.88	4.2	3.1
GPT-4	94.6	0.93	4.7	3.5

The results show that GPT-4 significantly outperforms all other models, achieving an accuracy of 94.6% and an F1-score of 0.93. This superior performance is a testament to its advanced architecture and its capacity to provide fluent and contextually appropriate responses.

4.2. Discussion: Interpreting Results Through a Theoretical Lens

The superior performance of GPT-4 is not merely a technical achievement; it can be interpreted through our theoretical framework, as the model's capabilities directly satisfy the key dimensions of service quality. GPT-4's superior performance across metrics of accuracy, fluency, coherence, and topical relevance compared to traditional methods can be attributed to its advanced architecture (OpenAI, 2023; Hagos et al., 2024) and the effectiveness of fine-tuning and prompt engineering (Vatsal & Dubey, n.d.; White et al., 2023; Reinhard et al., 2024). While traditional models like Naive Bayes and SVM are viable for general sentiment analysis, they struggle with subtle emotional expressions and lack specificity for educational contexts (Guia et al., 2019; Ranoliya et al., 2017; Ahmad et al., 2021).

4.3. GAI's Alignment with Service Quality Dimensions

The high accuracy (94.6%) and F1-score (0.93) of GPT-4 directly align with the **Assurance** dimension of the SERVQUAL model. The model's ability to consistently generate factually correct and contextually relevant responses conveys a sense of knowledge and trustworthiness, which is crucial for building user confidence in an automated system. The significant advantage of GAI models over traditional classifiers lies in their capacity to generate dynamic, context-aware responses, thereby fulfilling the **Responsiveness** dimension of service quality more effectively than static, rule-based systems.

4.4. The Role of Empathy in Automated Systems

The qualitative feedback, gathered from 10 human evaluators, further reinforces GAI's potential. GPT-4 received particularly high scores for **Empathy**, with evaluators appreciating its professional and approachable tone. For example, when confronted with an emotionally charged query, such as "I'm feeling overwhelmed with course registration, what should I do?", GPT-4 provided supportive language and actionable steps. This ability to respond with sensitivity and understanding represents a major step forward, as it moves automated customer service from a purely transactional function to a more personalized and supportive interaction.

4.5. Challenges and Ethical Considerations

Despite its high performance, GAI models, including GPT-4, are not without their challenges. The outputs sometimes included hallucinations—the generation of incorrect or irrelevant information—and were occasionally overly verbose. Furthermore, biases present in the training data occasionally resulted in culturally mismatched or non-inclusive responses. These findings underscore the need for careful content filtering, custom training, and human oversight in the deployment of GAI systems.

5. Conclusion and Future Work

5.1. Conclusion

This research work provides a comparative evaluation of several AI approaches in the context of university customer service. The findings demonstrate that Generative AI models, particularly GPT-4, represent a significant advancement over traditional methods, delivering more optimal and personalized student support. The findings demonstrate that Generative AI models, particularly GPT-4, represent a significant advancement over traditional methods, delivering more optimal and personalized student support (OpenAI, 2023; Hagos et al., 2024; Odede & Frommholz, 2024). GPT-4's superior performance across metrics of accuracy, fluency, coherence, and topical relevance can be attributed to its advanced architecture (OpenAI, 2023; Hagos et al., 2024) and the effectiveness of fine-tuning and prompt engineering (Vatsal & Dubey, n.d.; White et al., 2023; Reinhard et al., 2024). This success can be theoretically explained as the model's ability to excel on key service quality dimensions, including assurance, responsiveness, and empathy. However, the study also reveals that automated systems without human intervention still carry risks, especially when dealing with sensitive or critical student needs. While Generative AI holds great potential, its integration must be carefully managed to ensure it enhances, rather than hinders, the student experience by addressing concerns of privacy, bias, and the need for continuous human oversight.

The integrity of AI deployment hinges on robust ethical compliance measures, especially concerning factual accuracy and safety (Khennouche et al., 2024; Nicolescu & Tudorache, 2022; Paudel & Acharya, 2024; OpenAI, 2023; Dwivedi et al., 2023). Ethical compliance (screening for hallucinations and inappropriate language) was measured through manual review of a subset of outputs to verify factual correctness and absence of harmful or biased content (OpenAI, 2023).

5.2. Limitations

This study offers important insights into the application of generative AI and machine learning models; at the same time, several areas present opportunities for refinement in future research. These aspects, which were beyond the ethical and practical scope of the present work, are acknowledged here to guide subsequent investigations.

First, while the dataset used in this study (N=200) provided a meaningful starting point, future work could benefit from larger and more systematically sampled datasets to enhance representativeness and statistical power. Similarly, incorporating inferential statistical analyses alongside descriptive measures would enable a more precise assessment of model performance and provide stronger evidence for comparative claims.

The human evaluation component of this study yielded valuable qualitative feedback; however, expanding the evaluator pool and adopting standardized protocols would further strengthen the reliability and generalizability of these insights. Finally, future studies may consider implementing more controlled experimental designs, particularly in standardizing computational resources, to ensure that performance comparisons across models reflect intrinsic capabilities rather than external factors.

By addressing these areas, future research can build on the foundation laid by this study, deepening the understanding of generative AI's potential in educational contexts and advancing toward more robust and generalizable findings.

5.3. Future Work

The limitations of this exploratory study provide a clear and actionable roadmap for future research. A primary area for future work lies in the ethical and methodological advancements needed to scale GAI for university customer service. This includes specialized applications like predictive modeling of faculty engagement, which can be enhanced using deep learning models combined with data augmentation techniques (such as synthetic minority oversampling, categorical permutation, and noise injection) to address data scarcity common in educational contexts (Yahmadi et al., 2025; Fang et al., 2022; Dhatterwal et al., 2023).

First, to address the issue of model hallucinations and factual inaccuracy, future research could incorporate Retrieval-Augmented Generation (RAG) (Khennouche et al., 2024; Shuster et al., 2021). This approach would allow models to dynamically retrieve the most current and authoritative information from a university's knowledge base, thereby mitigating the problem of LLMs providing plausible-sounding but factually incorrect answers (Khennouche et al., 2024; Cai et al., 2022; Odede & Frommholz, 2024). RAG enables the model to utilize contextually relevant and up-to-date embeddings from a vector database rather than relying solely on embedded knowledge learned during pre-training (Khennouche et al., 2024; Odede & Frommholz, 2024).

Another area for improvement is the incorporation of user feedback loops and Reinforcement Learning from Human Preferences (RLHF) (Ouyang et al., 2022). By using feedback from students and staff, the AI system can continuously learn from real-world interactions and better align its responses with student expectations, improving both its accuracy and empathy (Ouyang et al., 2022). Future research should also explore extending the framework to encompass multilingual capabilities, catering to a variety of cultural and linguistic needs (Khennouche et al., 2024).

References

Ahmad, A., Al-Dmour, R., Alkhazaleh, M., Al-Fazza, R., Al-Shaboul, Y., & Al-Tahat, S. (2021). AI customer service system with pre-trained language and response-ranking models for university admissions. In *2021 IEEE International Conference on Research in Informatics and Systems (IRI)*. IEEE.

Alkishri, W., Alharthi, F., & Alkhadher, M. (2025a). Leveraging chatbots for student support and learning outcomes: A systematic literature review and research agenda. *Journal of Logistics, Informatics and Service Science, 12*(3), 264–282.

Alkishri, W., Yousif, J. H., Al Husaini, Y. N., & Al-Bahri, M. (2025b). Conversational AI in education: A general review of chatbot technologies and challenges. *Journal of Logistics, Informatics and Service Science, 12*(3), 264–282.

Cai, D., Wang, Y., Liu, L., & Shi, S. (2022). Recent advances in retrieval-augmented text generation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 3417–3419). ACM.

Carbonell, J. R. (1970). AI in CAI: An artificial-intelligence approach to computer-assisted instruction. *IEEE Transactions on Man-Machine Systems, 11*(4), 190–202.

Chen, C., Zhang, X., Ju, S., Fu, C., Tang, C., Zhou, J., & Li, X. (2019). AntProphet: an intention mining system behind Alipay's intelligent customer service bot. *IJCAI*, 6497–6499.

Dhatterwal, J. S., Baliyan, A., & Prakash, O. (2023). Reliability Driven and Dynamic Re-synthesis of Error Recovery in Cyber-Physical Biochips. In *Cyber Physical Systems* (pp. 15–34). Chapman and Hall/CRC.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., & Wright, R. (2023). So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, Article 102642.

Fang, J., Tang, C., Cui, Q., Zhu, F., Li, L., Zhou, J., & Zhu, W. (2022). Semi-supervised learning with data augmentation for tabular data. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management* (pp. 3928–3932). ACM.

Følstad, A., & Brandtzaeg, P. B. (2020). Users' experience with chatbots: Findings from a questionnaire study. *Qualitative User Experience, 5*(3).

Følstad, A., & Taylor, C. (2021). Investigating the user experience of customer service chatbot interaction: A framework for qualitative analysis of chatbot dialogues. *Qualitative User Experience, 6*(6).

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM, 63*(11), 139–144.

Guia, M., Silva, R., & Bernardino, J. (2019). Comparison of Naïve Bayes, Support Vector Machine, Decision Trees and Random Forest on Sentiment Analysis. In *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2019)*, 525–531.

Gupta, P., Ding, B., Guan, C., & Ding, D. (2024). Generative AI: A systematic review using topic modelling techniques. *Data and Information Management, 8*, Article 100066.

Hagos, D. H., Tadesse, M. M., Gidey, E. F., & Abdu, J. M. (2024). Recent advances in Generative AI and large-language models: Current status, challenges, and perspectives. *IEEE Transactions on Artificial Intelligence*.

Jensen, M., Restrepo, M., Sands, S., Ferraro, C., & Campbell, C. (2024). The paradoxes of generative AI-enabled customer service: A guide for managers. *Business Horizons, 67*(2), 215–225.

Ji, M., & Zhang, X. (2022). Research on semantic similarity calculation methods in Chinese financial intelligent customer service. *International Journal of Computer Applications in Technology, 68*(2), 143–149.

Khennouche, F., Elmir, Y., Himeur, Y., Djebari, N., & Amira, A. (2024). Revolutionizing generative pre-traineds: Insights and challenges in deploying ChatGPT and generative chatbots for FAQs. *Expert Systems With Applications, 246*, Article 123224.

Kitchenham, B. (2004). *Procedures for performing systematic reviews: vol. 33*, (pp. 1–26). Keele University.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.

Li, C., Zhang, X., & Chen, G. (2022). Construction of Intelligent Customer Service System on E-Commerce Platform and Its Impact on User Experience. In *Proceedings of the 2022 International Conference on Information Systems and Service Science (ICISS)*. ACM.

Misischia, C. V., Poecze, F., & Strauss, C. (2022). Chatbots in customer service: Their relevance and impact on service quality. *Procedia Computer Science, 201*, 421–428.

Nicolescu, L., & Tudorache, M. T. (2022). Human-computer interaction in customer service: The experience with AI chatbots—A systematic literature review. *Electronics, 11*(10), 1579.

Odede, J., & Frommholz, I. (2024). JayBot – Aiding University Students and Admission with an LLM-based Chatbot. In *Proceedings of the 2024 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '24)*. ACM.

Okoli, C. (2015). A guide to conducting a standalone systematic literature review. *Communications of the Association for Information Systems, 37*, 879–910.

OpenAI. (2023). *GPT-4 Technical Report*.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., et al. (2022). Training language models to follow instructions with human feedback. *arXiv preprint arXiv:2203.02155*.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71.

Paudel, S. R., & Acharya, N. (2024). Factors affecting behavioral intention to use ChatGPT: Mediating role of attitude. *Journal of Service, Innovation and Sustainable Development, 5*(2), 143–162.

Ranoliya, B. R., Raghuwanshi, N. S., & Singh, S. (2017). An AIML based chatbot for education domain using LSA. In *2017 international conference on computer, communication and signal processing (ICCCSP)* (pp. 1–6). IEEE.

Reinhard, P., Li, M. M., Peters, C., & Leimeister, J. M. (2024). Generative AI in customer support services: A framework for augmenting the routines of frontline service employees. *Proceedings of the 57th Hawaii International Conference on System Sciences (HICSS)*.

Sahoo, D., Maity, M., & Misra, S. (2024). A survey of prompting techniques for large language models. *arXiv preprint arXiv:2401.07119*.

Shawar, B. A., & Atwell, E. (2007). Chatbots: are they really useful? *Journal for Language Technology and Computational Linguistics, 22*(1), 29–49.

Shuster, K., Poff, S., Chen, M., Kiela, D., & Weston, J. (2021). Retrieval augmentation reduces hallucination in conversation. *Findings of the Association for Computational Linguistics: EMNLP 2021*, 3784–3803.

Tan, Y. H., Chua, H. N., Low, Y.-C., & Jasser, M. B. (2024). Current landscape of Generative AI: Models, applications, regulations and challenges. *Proceedings of the 2024 IEEE Conference on Cyber Security and Cloud Engineering (ICCSCE)*. IEEE.

Vatsal, S., & Dubey, H. (n.d.). *A survey of prompt engineering methods in large language models for different NLP tasks*.

vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., & Cleven, A. (2015). Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research. *Communications of the Association for Information Systems, 37*.

Wang, Y. (2018). *Intelligent customer service system design based on natural language processing*.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly, 26*, xiii–xxiii.

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). *A prompt pattern catalog to enhance prompt engineering with ChatGPT*. ArXiv preprint arXiv:2302.11382.

Yahmadi, H., Al-Shammakhi, K., & Al Husaini, Y. (2025). Predicting faculty engagement in communities of practice: A neural network approach with synthetic data augmentation. *Journal of Logistics, Informatics and Service Science, 12*(3), 141–162.