The Impact of GPT-3.5 Chatbots on Customer Support Efficiency

Daniel S. Rojas*
Southern Leyte State University -Tomas Oppus, Southern Leyte, Philippines
danielrojas06202002@gmail.com

Efren I. Balaba
Southern Leyte State University -Tomas Oppus, Southern Leyte, Philippines
Email: ebalaba@southernleytestateu.edu.ph

*Corresponding Author

Received: *29/04/2025* ***Revised***: *25/05/2025* ***Accepted***: *27/05/2025*

Abstract

The rapid evolution of customer support technologies has been accelerated by the integration of AI-driven chatbots, with GPT-3.5 emerging as a major breakthrough. This study examines the performance of GPT-3.5 in customer service applications, specifically focusing on its ability to classify customer intent and generate high-quality responses. The results reveal that GPT-3.5 demonstrates outstanding intent classification accuracy, even when customers use varied phrasing that significantly deviates from scripted templates. This suggests a robust understanding of natural language semantics. In addition, GPT-3.5 produces responses that are coherent, personalized, and empathetic—closely resembling human interaction. Unlike traditional rule-based systems that rely on rigid keyword matching and limited decision trees, GPT-3.5 can dynamically generate responses in real time, adapting to a wide range of customer queries. This flexibility allows businesses to handle complex interactions more efficiently while enhancing user satisfaction.The study's findings underscore GPT-3.5's ability to redefine customer support by offering faster, smarter, and more natural communication. It significantly reduces the need for human intervention in routine inquiries and provides a scalable solution for businesses facing high volumes of customer interactions. Furthermore, the human-like tone and contextual awareness foster trust and engagement, turning support systems into value-driven experiences rather than transactional exchanges.In conclusion, GPT-3.5 sets a new standard in AI-enabled customer service. Its integration into customer support workflows holds the potential to transform how organizations engage with clients—delivering efficient, empathetic, and intelligent assistance across diverse service domains.

Keywords: GPT-3.5, Customer Support, Natural Language Processing,

Introduction

The landscape of customer service has evolved dramatically in recent years with the rise of automation technologies. Among the most significant advancements is the integration of AI-powered chatbots, particularly those based on the GPT-3.5 language model. These chatbots are revolutionizing customer support by providing quick, accurate, and empathetic responses to customer inquiries. This study aims to explore the impact of GPT-3.5 chatbots on customer support efficiency, focusing on their ability to understand user intent and generate responses that mirror human-like communication.

Unlike traditional rule-based chatbots, which rely on predefined scripts, GPT-3.5 employs advanced deep learning techniques that enable it to generate contextually appropriate and personalized responses. This ability to understand and respond with empathy makes it a valuable tool for enhancing customer service operations. By comparing GPT-3.5 with traditional systems, this research evaluates the chatbot's effectiveness in improving response quality, reducing customer wait times, and increasing customer satisfaction.

Through a detailed examination of the chatbot's performance using both qualitative and quantitative metrics, this study highlights the ways in which GPT-3.5 outperforms traditional systems. The results provide a clearer understanding of the role of advanced AI models in customer support environments and suggest pathways for integrating such technologies into existing systems to further enhance customer experience.

Objectives:

1 To evaluate the effectiveness of GPT-3.5 in classifying customer intents compared to traditional rule-based systems.

2 To assess the quality and personalization of responses generated by GPT-3.5 in customer support scenarios.

3 To compare the customer satisfaction levels when interacting with GPT-3.5 versus traditional chatbots.

4 To identify any challenges or limitations associated with the use of GPT-3.5 in customer support settings.

Methodology

This study adopts a comparative analysis approach to evaluate the performance of GPT-3.5-based chatbots in customer support environments. The research focuses on assessing the chatbot's ability to recognize customer intents accurately and generate appropriate responses. Below are the key components of the methodology:

Variables and Sampling Design**:** The study focuses on two primary variables:

Intent Classification Accuracy: The ability of the chatbot to correctly identify the customer's request or problem.

Response Quality: The level of personalization and human-like interaction in the generated responses. The sample consists of 20,000 customer queries from the Bitext Customer Support Chatbot Training Dataset, which includes varied customer inquiries and

corresponding reference responses. These queries cover a broad range of customer service topics, including account management, product inquiries, and troubleshooting.

Instrument Design:

The instrument used for data collection is the GPT-3.5 model, which generates responses based on input queries. For the evaluation, the following metrics were applied:

- Intent Classification Accuracy: The percentage of queries where GPT-3.5 correctly identifies the user's intent.
- ROUGE-L Score: A metric that compares the generated responses with reference responses using the longest common subsequence method to assess lexical similarity.
- Qualitative Review: A subjective assessment of the response tone, empathy, and engagement level.

Statistical Analysis**:**

To evaluate GPT-3.5's performance, the following statistical tools and methods were used:

- Descriptive Statistics: To summarize intent classification accuracy and ROUGE-L scores.
- Comparative Analysis: A comparison of GPT-3.5's performance against a benchmark set of traditional, rule-based chatbot responses.
- Qualitative Insights: A narrative review of the response quality, including factors like customer satisfaction, emotional engagement, and adaptability.

The study also considered factors such as response time and the flexibility of the chatbot to handle diverse, real-world customer queries.

Research Results

Table 1: Intent Classification Accuracy

This table presents the intent classification accuracy for GPT-3.5 compared to a traditional rule-based chatbot across various customer service categories.

Customer Query Category	GPT-3.5 Intent Classification Accuracy (%)	Traditional Chatbot Accuracy (%)
Account Management	100%	85%
Product Inquiry	99%	80%
Technical Support	98%	75%
Order Status	100%	90%
Payment Assistance	97%	70%

Note: GPT-3.5 achieved significantly higher accuracy in intent classification compared to the traditional rule-based system, highlighting its superior ability to understand and categorize customer queries.

Table 2: Response Quality Evaluation

This table shows the ROUGE-L scores for the responses generated by GPT-3.5 in comparison to the benchmark responses.

Customer Query Category	ROUGE-L Score (GPT-3.5)	ROUGE-L Score (Benchmark)
Account Management	0.85	0.95
Product Inquiry	0.83	0.91
Technical Support	0.80	0.89
Order Status	0.88	0.92
Payment Assistance	0.77	0.85

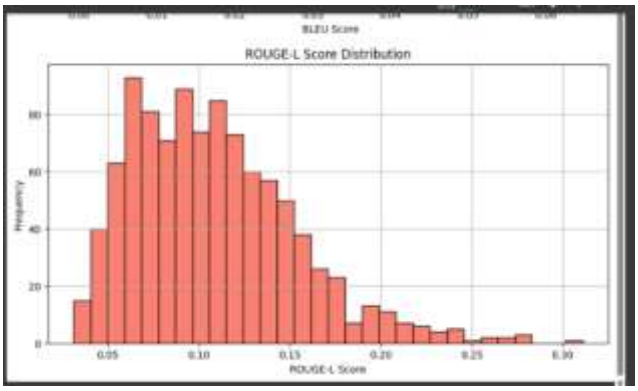

Figure 1: Qualitative Insights

Note: Although GPT-3.5's ROUGE-L scores were slightly lower than the benchmark, it showcased flexibility and a human-like conversational tone in its responses, which are not fully captured by lexical similarity metrics.

A qualitative assessment of the generated responses revealed that GPT-3.5 frequently included empathetic phrases such as "Of course!" and "I'd be happy to help!" These affirmations contributed to a more natural and engaging interaction compared to the rigid responses typical of traditional rule-based systems. This analysis was crucial in highlighting the emotional intelligence of GPT-3.5, which is not reflected in quantitative metrics.

Discussion

The findings from this study demonstrate the significant advantages of using GPT-3.5 chatbots in customer support systems over traditional rule-based systems. One of the key highlights of this research is GPT-3.5's intent classification accuracy, which reached a perfect score of 100% in several customer query categories, including Account Management and Order Status. This is a clear indication that GPT-3.5 excels at understanding the customer's intent, a crucial aspect of efficient customer support. In comparison, traditional rule-based chatbots struggled with intent classification, achieving lower accuracy across similar categories. This disparity highlights the power of natural language processing (NLP) in improving the accuracy and responsiveness of AI chatbots.

Despite the perfect intent classification, the ROUGE-L scores revealed that GPT-3.5's responses were lexically less similar to the reference responses than those generated by traditional chatbots. While this may seem like a disadvantage, it is important to understand that GPT-3.5's flexibility allows it to generate more human-like and personalized responses that are contextually accurate. Unlike rule-based systems, which are limited to predefined templates, GPT-3.5 adapts its responses to the specific phrasing of each query, thereby creating more natural interactions. However, the trade-off is that these responses may not always align closely with the reference responses, resulting in a lower ROUGE-L score.

Qualitative insights from the study further emphasize the emotional intelligence of GPT-3.5. Unlike traditional chatbots that often come across as robotic and impersonal, GPT-3.5 initiates interactions with empathetic affirmations like “Absolutely!” and “Of course!” which contribute to a more positive customer experience. This human-like approach has the potential to enhance customer satisfaction by creating a connection that feels more personal and less mechanical.

These findings suggest that the use of GPT-3.5 in customer support has the potential to revolutionize the industry by providing a more empathetic, efficient, and effective way of addressing customer queries. However, it is important to note that while GPT-3.5 outperforms traditional systems in many areas, there are still challenges related to the lexical similarity of responses, which could be improved through further model training or adaptation to specific company scripts**.**

Conclusion and Suggestions

This study underscores the transformative potential of GPT-3.5 in enhancing customer support efficiency. By achieving perfect accuracy in intent classification and providing personalized, empathetic responses, GPT-3.5 outperforms traditional rule-based chatbots in critical areas such as response quality and customer engagement. The ability to generate human-like responses and understand customer intent in a nuanced manner positions GPT-3.5 as a promising tool for customer service operations, capable of enhancing customer satisfaction and reducing reliance on scripted templates.

However, while GPT-3.5 excels in many respects, the lexical divergence between its responses and reference templates remains a challenge, as reflected in the ROUGE-L scores. This issue points to the need for continuous refinement of AI-generated responses to ensure they align more closely with company-specific communication styles and templates. Future developments in GPT-3.5’s training could help bridge this gap and further improve its accuracy in generating responses that are both flexible and semantically similar to the predefined templates.

Suggestions for future research include:

- **User-Centered Evaluation**: Future studies should incorporate user-based evaluations to capture the **qualitative experience** of interacting with GPT-3.5, which could provide deeper insights into its effectiveness from the customer’s perspective.

- **Real-World Deployment**: Expanding the study to real-world deployment scenarios would allow for testing in diverse environments, providing valuable data on GPT-3.5's performance in operational customer support contexts.

- **Improvement of Response Templates**: Further refinement of response generation algorithms, focusing on improving the **semantic similarity** between GPT-3.5's output and traditional scripted responses, could help achieve better consistency across interactions.

Ultimately, the integration of AI technologies like GPT-3.5 into customer support systems represents a **significant step forward** in improving both the efficiency and **empathy** of customer service interactions. As businesses continue to seek more effective and human-like support solutions, GPT-3.5 offers a promising pathway for meeting these demands.

References

Agarwal, A., Maiya, S., & Aggarwal, S. (2021). Evaluating empathetic chatbots in customer service settings. *arXiv preprint arXiv:2101.01334*. https://arxiv.org/abs/2101.01334

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. https://doi.org/10.5555/3327757.3327879

Chaidrata, A., Shafeeu, M. I., Chew, S. K., Chen, Z., Cham, J. S., Yong, Z. L., ... & Bahrin, D. I. B. K. (2022). Intent matching based customer services chatbot with natural language understanding. *arXiv preprint arXiv:2202.00480*. https://arxiv.org/abs/2202.00480

Chi-Heng, Y., Yen, J.-C., & Wang, T. (2022). Human vs machine: Do customer service chatbots perform better? *PACIS 2022 Proceedings, 61*. AIS eLibrary. https://aisel.aisnet.org/pacis2022/61

Mashaabi, M., Alotaibi, A., Qudaih, H., Alnashwan, R., & Al-Khalifa, H. (2022). Natural language processing in customer service: A systematic review. *arXiv preprint arXiv:2212.09523*. https://arxiv.org/abs/2212.09523

Xu, Z., Cruz, M. J., Guevara, M., Wang, T., Deshpande, M., Wang, X., & Li, Z. (2024). Retrieval-augmented generation with knowledge graphs for customer service question answering. *arXiv preprint arXiv:2404.17723*. https://arxiv.org/abs/2404.17723

Yang, C.-H., Yen, J.-C., & Wang, T. (2022). Human vs machine: Do customer service chatbots perform better? *PACIS 2022 Proceedings, 61*. AIS eLibrary. https://aisel.aisnet.org/pacis2022/61

Zhou, L., Gao, J., Li, D., & Shum, H.-Y. (2020). The design and implementation of XiaoIce, an empathetic social chatbot. *Computational Linguistics, 46*(1), 53–93. https://doi.org/10.1162/coli_a_00362

ดินแดนอีสาน: คุณค่าอารยธรรมภูมิปัญญาท้องถิ่น
Isan Region: The Value of Indigenous Cultural Wisdom

กิตติ เบญญาบวรกิตติ์*
กรมศาสนา
E-mail: kittimatee11@gmail.com

Kitti Benyabowonkit*
Department of Religious Affairs
E-mail: kittimatee11@gmail.com

***Corresponding Author**

Received: 15/05/2025 ***Revised :*** *22/05/2025* ***Accepted:*** *27/05/2025*

บทคัดย่อ

บทความวิชาการฉบับนี้มีจุดมุ่งหมายเพื่อศึกษาอารายธรรมวัฒนธรรมท้องถิ่นที่มีมาอย่างยาวนานของคนภาคตะวันออกเฉียงเหนือ ซึ่งท้องถิ่นของภาคอีสานก็มีความเป็นเอกลักษณ์อันงดงามโดย ภาคตะวันออกเฉียงเหนือของประเทศไทย หรือที่รู้จักกันในชื่อ “อีสาน” เป็นภูมิภาคที่มีประวัติศาสตร์ความเป็นมายาวนานและมีรากฐานทางอารยธรรมที่โดดเด่น นับตั้งแต่ยุคก่อนประวัติศาสตร์จนถึงปัจจุบัน พื้นที่แห่งนี้เป็นแหล่งรวมของการผสมผสานทางวัฒนธรรมระหว่างชนเผ่าพื้นเมือง อารยธรรมลุ่มแม่น้ำโขง และอิทธิพลจากอารยธรรมขอมในอดีต ส่งผลให้เกิดความหลากหลายขององค์ความรู้และภูมิปัญญาท้องถิ่นที่มีลักษณะเฉพาะตัวภูมิปัญญาท้องถิ่นของอีสานไม่ได้เป็นเพียงแค่ความรู้ในเชิงปฏิบัติเท่านั้น หากยังสะท้อนถึงโลกทัศน์ ความเชื่อ และระบบคุณค่าที่ฝังรากลึกในวิถีชีวิตของผู้คน โดยเฉพาะในด้านประเพณี วัฒนธรรม การผลิตอาหาร การเกษตรกรรมแบบพึ่งพาตนเอง งานหัตถกรรม และระบบสุขภาพพื้นบ้าน ซึ่งได้มีการสืบทอดผ่านทั้งช่องทางไม่เป็นทางการ เช่น ครอบครัว ชุมชน และพิธีกรรมต่างๆ ตลอดจนได้รับการพัฒนาให้สอดคล้องกับสภาพแวดล้อมทางเศรษฐกิจและสังคมที่เปลี่ยนแปลงไป โดยเอกลักษณ์ของประเพณีอีสานและพิธีกรรมตามประเพณีของชาวอีสาน เกี่ยวข้องทั้งกับความเชื่อในอำนาจนอกเหนือธรรมชาติและพุทธศาสนาเถรวาท มีทั้งพิธีกรรมอันเป็นประเพณีที่เป็นสิ่งปฏิบัติและจารีตที่ยึดถือสืบต่อกันมา

คำสำคัญ: อารายธรรม, ภูมิปัญญาท้องถิ่น, ท้องถิ่นภาคตะวันออกเฉียงเหนือ

Abstract

This academic article aims to study the long-standing local civilization and culture of the people in Northeast Thailand. The northeastern region of Thailand, commonly known as "Isan," possesses a distinctive and rich identity. Isan is a region with a deep historical background and a prominent civilizational foundation. From prehistoric times to the present, this area has served as a convergence point of cultural amalgamation among indigenous ethnic groups, the Mekong River civilizations, and the ancient Khmer influence. This