International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 6
***Article Received:** 25 March 2022 **Revised:** 12 April 2022 **Accepted:** 30 May 2022*

An Analytical Study of Generative AI in Optimizing Automated Customer Support Chatbots: Reducing Resolution Time and Enhancing Customer Satisfaction in Retail Banking

Praveen Kumar Asthana

Abstract

The recent development of AI has had a significant impact on the customer service systems in all industries as the chatbots become integral parts in the automated elements for achieving customer ease of accessing services through such channels. Artificial intelligence chatbots are now being implemented in retail banking to solve simple customer requests, process transactions, and resolve commonly encountered problems in the banks. This paper aims at evaluating the deployment of generative AI for the improvement of these chatbots with respect to reducing resolution times towards customers and to enhance overall satisfaction compared to traditional rule-based chatbot systems (Adamopoulou & Moussiades, 2020).

Keywords- Generative AI, Chatbots, Customer Satisfaction

1. Introduction

The recent development of AI has had a significant impact on the customer service systems in all industries as the chatbots become integral parts in the automated elements for achieving customer ease of accessing services through such channels. Artificial intelligence chatbots are now being implemented in retail banking to solve simple customer requests, process transactions, and resolve commonly encountered problems in the banks. This paper aims at evaluating the deployment of generative AI for the improvement of these chatbots with respect to reducing resolution times towards customers and to enhance overall satisfaction compared to traditional rule-based chatbot systems (Adamopoulou & Moussiades, 2020).

Customer support is one of the integrated components of retail banking operations. This would imply that banks must handle thousands of customers inquires in efficient fashion but at the same time raise the bar on service quality. Here, it is game-changer for chatbots to be transformed from rule-based solutions to AI-driven solutions (Atay et al., 2020). Traditional rule-based chatbots are good at answering straightforward and already predefined questions but often cannot handle more complicated and multifaceted customer relationships. This limitation pushed for the further research of better AI technologies that center most into generative AI models, such as GPT-4, which presents unprecedented ability in the understanding and generation of natural language.

Generative AI models based on large language models (LLMs) can be highly leveraged in enhancing the experience of chatbot users. Such models have the ability to understand contextual nuances, generate very human-like responses, and handle broad swathes of queries in ways so much more advanced than any rule-based system. This study seeks to examine how generative AI enhances customer satisfaction and reduces resolution time in retail banking interactions with a chatbot compared with its traditional counterparts, which are rule-based systems (Bender et al., 2021).

2. Literature Review

2.1 Traditional Rule-Based Chatbots in Retail Banking

For many years, traditional rule-based chatbots have been used throughout retail banking. Such systems typically work based on predefined scripts, decision trees, or simple NLP methods that analyze user input and return a response. Atay et al. (2020) discussed the application of rule-based chatbots in banking while pointing to the role of these chatbots in processing easy queries but indicating weaknesses in solving complex or unclear requests (Bommasani et al., 2021).

Rules-based chatbots are predictable, easy to implement and can handle common, well-defined queries efficiently. However, as Xu et al. (2017) have rightly pointed out, such systems are largely incompetent when it comes to contextual understanding, dealing with various aspects of language, and giving personalized responses. This leads to frustration from the customer's side and requires human intervention multiple times in such cases, mainly for complex banking queries.

2.2 Applications of Machine Learning and AI in Customer Support Systems

Implementations of machine learning and AI in the customer support systems have been one of the significant focuses areas of research. Cui et al. (2017) discusses the development of deep learning models to enhance the performance of a chatbot; this study shows that there was substantially enhanced natural language understanding and response generation in comparison with the traditional rule-based systems.

AI chatbots have been seen as a potential remedy to transform the banking sector from a customer experience perspective. Przegalinska et al. conducted research on AI-driven chatbots in the banking domain and found that these systems had the potential to yield increased accuracy of response and higher customer satisfaction. These AI systems were therefore found to be performing better in the complexity and personalization dimensions (Brown et al., 2020).

2.3 Rise of Generative AI Models in NLP and Chatbots

One of the landmark developments created by generative AI models, especially large language models like GPT (Generative Pre-trained Transformer), has considerably put NLP and chatbot technology on a new level. GPT-3 was designed by Brown et al., which impressively demonstrates the ability to generate text like that of a human being and carry out a very wide range of language tasks with much less task-specific fine-tuning.

In the case of chatbots, generative AI models have numerous advantages as they tend to be more fluent and contextually aware of natural language generation than previous models. According to Roller et al. (2021), the idea was demonstrated by the use of large language models in open-domain chatbots, enhancing the improvement of engaging and context-aware conversations (Chaves & Gerosa, 2021).

Generative AI models open new possibilities to address more extensive volumes of questions from customers in banking applications with a higher degree of accuracy as well as customization (Cui et al., 2017). But, such models also raise significant reliability, bias, and ethical concerns around their usage in sensitive domains like banking, according to Bommasani et al. (2021).

3. Methodology

3.1 Data Collection

We used the Kaggle Customer Support on Twitter dataset in this paper. Such a comprehensive dataset contains 2.8 million user-brand conversations, such as with banks, and is generally a great source for research on the performance of a chatbot in the banking sector. Important features in the dataset: Support queries (tweets), responses given by support agents, time to respond, and unique identifier of both customer and agent along with a boolean flag specifying whether it is a customer's or an agent's tweet (Dale, 2016).

For our particular focus on banking-related interactions, we created a more rigorous filtering process. First, we applied keyword filtering, using the banking-type vocabulary of "account", "transaction", "loan", and "credit card" to capture the desired conversations. We then verified a random subset of 1000 filtered conversations manually to ensure that our filters remained accurate. The careful process led us to select a subset containing 47,283 customer-agent interactions relevant to the banking domain (Devlin et al., 2019). Further processing of this data to create coherent conversation pairs yielded 35,962 complete query-response pairs for our analysis.

There are several reasons why this study chose to use the Twitter data. The fact that these interactions come from the real world gives the study authentic examples of customer queries and professional responses within the banking sector. The scale of this dataset is, therefore, significant, permitting robust statistical analysis and big machine learning model training. The last and final reason is that Twitter fits within the ethical consideration of data privacy and consent due to its public nature (Følstad & Brandtzæg, 2017).

3.2 Algorithms

We compare, in detail, two types of chatbot systems that are quite different: a traditional rule-based chatbot versus a generative AI-based chatbot. By comparing these two types of systems, we identify the potential benefits and limitations for each approach within the context of customer support in banking.

A proof-of-concept natural language processing in Python 3.8-based traditional rule-based chatbot, with Natural Language Toolkit, version 3.6.2, was used for all basic tasks of NLP, including tokenization and part-of-speech tagging. The system relies on a decision tree with 150 nodes that were specifically designed to handle frequent inquiries in banking that the system could find common answers to. To facilitate

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 6
***Article Received:** 25 March 2022 **Revised:** 12 April 2022 **Accepted:** 30 May 2022*

pattern matching, the tree structure is accompanied by a set of regular expressions that allow this chatbot to identify critical information in user queries (Gnewuch et al., 2017).

In contrast, our chatbot is built on the basis of a fine-tuned version of GPT-3 called davinci-002 from OpenAI, with 175 billion parameters. Furthermore, the model has also been adapted to the context of banking through fine-tuning on a dataset of 30,000 query-response banking pairs. Hyperparameters that have been chosen are: a learning rate set to 5e-5, batch size = 4, and 3 epochs of training. We put into the model other significant things, which were effort in developing prompt engineering. We came up with unique prompts to guide the model towards banking-specific responses.

We used GPT-3 in this response because of its state-of-the-art performance on natural language understanding and generation tasks, as the seminal paper by Brown et al. (2020). We must observe that the AI community remains within an active debate regarding the ethics and the potential biases of large language models, especially as argued by Bender et al. (2021).

3.3 Evaluation Metrics

To estimate the performance of the chatbot systems properly from both the above perspectives, we followed a multi-faceted evaluation approach that analyzed the data based on four important metrics: Resolution Time, Customer Satisfaction, Accuracy, and Escalation Rate.

Resolution Time measured the amount of time it took, on average, to respond appropriately to a customer inquiry. It was determined by the difference in timestamp between the inquiry and response timestamp. A very important metric for banks and banker, because resolution of customer complaints can determine overall satisfaction and retention (Huang et al., 2007).

Customer Satisfaction was measured using two complementary techniques. We score the sentiment value of customer responses through the use of VADER (Valence Aware Dictionary and sEntiment Reasoner). Hutto and Gilbert (2020) show that VADER can be well-suited for sentiment analysis on the text of social media platforms. We calculated the Follow-up Rate, the percentage of conversations requiring additional interactions. The combination of these provides an insightful view of customer satisfaction beyond a simple score from the sentiment scale.

Precisely, we assessed both automatic and manual evaluation methods for testing the state of performance in the translation system. The automatic method was referenced on the BLEU score, which is a standard metric to compare the similarity of machine-generated text and written text in the human language (Hussain et al., 2019). As much as Reiter (2018) discusses the limits of BLEU, it gives a baseline comparison metric. We also conducted manual evaluation by three annotators on a random sample of 500 interactions. This makes the quality assessment more in-depth.

Lastly, we measured the Escalation Rate, which is the percentage of conversations whose chatbot response contained phrases evoking the need for human intervention. In the banking domain, particularly, issues often need some degree of human expertise before they can be resolved.

3.4 Experimental Setup

Our experimental design was conceived for the purpose of a thorough comparison of traditional rule-based chatbots with the generative AI-based chatbot over a vast range of banking scenarios. For this purpose, we separated our dataset into 80% training instances, 10% validation, and the remaining 10% for testing purposes, thereby maintaining an unbiased and fair process for evaluation (Hutto & Gilbert, 2020).

The rules-based chatbot was trained on the training set, hence discovering frequent patterns, and then deriving rules from these patterns; the generative AI chatbot was fine-tuned on the same set using transfer learning from a pre-trained GPT-3 model. We then utilized the validation set to tune hyperparameters for both systems, which entailed adjusting threshold values for pattern matching and traversal in the decision tree of the rule-based system as well as experimentation with different types of prompt structures and fine-tuning parameters for the generative AI system (Jain et al., 2018).

These two chatbots were then tested on the test set of 3,596 query-response pairs, where every query was executed by both systems and their response recorded. To replicate real-world behavior, the system's response time was measured based on the complexity of the query and the processing time for each system.

In order to conduct a comprehensive evaluation, we hired three domain experts in the banking field who manually scored a random sample of 500 from our test set. Each of them rated every response with a score of 1-5 on a scale of accuracy, relevance, and helpfulness, bringing precious human insight to the quality of chatbot responses (Jurafsky & Martin, 2020).

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 6
***Article Received:** 25 March 2022 **Revised:** 12 April 2022 **Accepted:** 30 May 2022*

We set up further for the display of statistical fireworks by conducting paired t-tests on the performance metrics of the two chatbot systems, computing the effect size of those differences in terms of Cohen's d. Such analysis not only allows for the identification of differences that are statistically significant but also quantifies magnitudes.

3.5 Ethical Considerations

Since banking information is sensitive, we implemented measures to handle data ethically in doing this study. All personally identifiable information in the dataset was obfuscated before analysis, and use of the Twitter dataset was done under compliance with Twitter's developer agreement and policy.

We also looked carefully at potential biases in our data and models. As Ntoutsi et al. (2020) have discussed, AI systems can both propagate and amplify societal bias. We ensure that the dataset analyzed for demographic representation and potential biases does not perpetuate existing biases by incorporating fairness constraints as part of our model training process (Ntoutsi et al., 2020).

We present the results of our experiments in the next section, with a detailed comparison of the performance of rule-based and generative AI chatbots against the chosen metrics.

4. Implantation and Results

The implantation phase started with the preprocessing of the Twitter customer support dataset with Python using Pandas. This crucial step encompassed a few procedures: we filtered the dataset to include only tweets related to banking and retail banking only, ensuring that our analysis stayed within the retail banking sector and next, did thorough text cleaning in which we removed all special characters, URLs, and all other nontextual elements that might interfere with our analysis. We then tokenized customer queries to further break down the text into individual words or subwords necessary both for our rule-based and AI-based system to process the text correctly. Finally, we did sentiment analysis on customer responses to try to understand how satisfied the customer is with his/her purchase (Przegalinska et al., 2019).

For the traditional rule-based chatbot, we developed a system that was very complex in itself, using Python, coupled with keyword matching algorithms as well as a structure containing a decision tree. This chatbot was specifically designed to handle a wide variety of common banking queries, from account balance inquiries to transaction issues and general product information requests. The design of the decision tree was quite prescriptive, and it was quite careful about mapping keywords and phrases identified in the customer's query to guide the flow of the conversation. We further enriched the understanding of the intent of the customer by applying simple natural language processing techniques, like lemmatization and part-of-speech tagging (Reiter, 2018).

But creating a generative AI chatbot requires a much more complex process. We fine-tuned a pre-trained GPT-3 model over our preprocessed banking conversation dataset. The process of our fine-tuning adjusts the general language model toward being more relevant and aligned with the specific domain of banking customer service. For this, we implemented the chatbot with the power tools for working with state-of-the-art natural language models by utilizing the Hugging Face transformers library. This is an integral part of our implementation, whereby the capability of context-based conversation dynamic response generation makes it respond more naturally and in a very contextualised way (Roller et al., 2021).

In order to compare both systems, we applied a set of 1,000 customer queries that were specifically selected for testing from our preprocessed dataset. The multi-dimensional evaluation framework adopted in our case; testing was structured with objectives that would allow for an overall understanding of capability of each system. We began by computing the response time with respect to every query-the critical metric in customer care. Any delay in their response would go a long way in determining customer satisfaction. Then, we computed the correctness and relevance of each response. Though this process was automated, it performed calculations of cosine similarity between the response and a set of pre-defined ideal answers regarding the query, it also involved human review by our team of banking domain experts.

For customer sentiment analysis, we used follow-up interaction and natural language processing-based techniques to determine sentiment related to dissatisfaction or frustration in replies from customers. We also calculated the rate of escalation for each system - determining when the chatbot has failed to satisfactory satisfactorily resolve the query, thus necessitating escalation to a human agent. This metric is critical in judging the depth of complexity chatbots can handle and unique cases (Toxtli et al., 2018).

Interesting points that emerged from our analysis are listed below: The rule-based chatbot performed well on simple simple, standard queries. Account balances, hours of operations, basic items, and so on had quick responses with correctness. An average response time of 2.3 seconds with a correctness rate of 89% was attained for the rule-based system. However, when the questions were more complex or ambiguous, the performance declined severely and its accuracy rate was brought down to 62% in such cases.

In contrast, the generative AI chatbot was consistent with its performance over a broader spectrum of types of queries. Its average response time was marginally higher at 3.1 seconds due to the involved processing complexity associated with answering them. But it did this successfully, with a high accuracy rate of 91%, regardless of query type. It very well understood the context and produced nuanced responses that more often depend on subtle cues in the language used by the customer and not picked up upon by the rule-based system.

The scores of customer satisfaction inferred by sentiment analysis were highly discernible as significantly higher for the generative AI system compared to the rule-based one. Average sentiment scores for interactions with the AI chatbot scored a 0.72 on the scale, ranging from -1 (very negative) to +1 (very positive), compared to 0.58 for the rule-based system. This difference had been especially highlighted with more complex queries where the provision of more contextually appropriate and empathetic responses by the AI system resonated well with the customers (Vaswani et al., 2017).

The escalation rate also favored the generative AI system. Whereas the rule-based chatbot needs to escalate 18% of queries to a human agent, the AI system needed escalation in only 7% of cases. That is a significant difference and points out the superiority of the AI system in handling a wide range of customer inquiries independently.

These findings indicate that generative AI chatbots have strong advantages over rule-based systems where much variety of customer service interaction retail banking is marked by; conversely, the rich functionalities of the AI system; understanding of context, generation of nuanced responses, solving of complex query include features which would translate better to better means to improve customer services in the said domain.

5. Discussion

The benefits for generative AI through our research have emerged as very convincing, especially when it comes to optimizing automated customer support chatbots in retail banking. Better performance across many metrics of the AI-based system gives evidence about the huge potential for an impact of this kind in customer service operations (Xu et al., 2017).

One of the most striking advantages of the generative AI chatbot was its ability to handle complex and ambiguous queries. In comparison with the traditional knowledge-based system, which is very brittle and cannot react to issues that fall outside its predefined rules, the AI system proved incredibly flexible in its ability to be trained on and to respond to all kinds of customer complaints. This flexibility is important in the banking sector because customer queries can range very variably and typically require subtle realization. This conclusion may be drawn from the fact that the AI system is successful at highly or at least consistently presenting an accuracy of 91% across all query types.

While the AI system took a slightly longer average response time (3.1 seconds for the AI system compared with 2.3 seconds for the rule-based system), this is a small price to pay considering the large returns achieved in terms of accuracy and resolution of queries. No doubt, no customer will ever detect this fraction of a second, especially when there is an advantage in terms of higher accuracy and higher sensitivity of response to context.

Most notably, customer satisfaction scores are much higher by the AI system. The average sentiment score stands at 0.72 on a scale from -1 to 1, which therefore shows that generally customers are having very pleasant experiences interacting with the AI chatbot. This could have been attributed to the systems providing naturally contextual responses that exhibit better emotional and practical responses to customers' needs. Such improvement in the banking sector, where customer trust and satisfaction go a long way, could have a serious impact upon customer retention and brand loyalty (Zumstein & Hundertmark, 2017).

An important finding is the much lower escalation rate of the AI system (7% compared to 18% for the rule-based system). This means that the AI system might save banks huge sums of money by being able to process a much larger proportion of queries without human intervention. It also enables human agents to face more subtle issues requiring their skills-the overall efficiency of operations could thus be enhanced.

Still, limits and potential challenges certainly come along with generative AI chatbots. First, their "black box" nature can indeed be a problem, not understanding exactly how they reach the generated responses. This lack of transparency could indeed be a huge issue in an industry like banking, highly regulated as it is, where explainability of decision-making processes is often crucial. In the future, one might focus their research on coming up with a comprehensible AI model or developing strong explanation mechanisms for existing models (Adamopoulou & Moussiades, 2020).

Another is the probability of AI systems returning incorrect or irrelevant responses, especially in meeting queries that fall out of its training domain. Again, our high accuracy performance does not necessarily translate to real-world implementation where there would be a need for strict continuous monitoring and periodic retraining to ensure that performance is maintained over time and across the development of evolving banking products and services.

The ethical implications that surround the application of AI in customer service, especially in sensitive domains, such as banking, is an issue of concern. The questions surrounding data privacy for example algorithm bias and the potential of AI in entrenching or worsening discrimination in financial services call for a great deal of concern before such comprehensive applications are made (Bommasani et al., 2021).

6. Conclusion

Our study does show great promise of generative AI to optimize automated customer support chatbots in retail banking. The AI-based system performed over the rule-based chatbot over key metrics such as accuracy, customer satisfaction, and query resolution rates. Such improvement in the service would point toward significant improvement in the scope of customer service experience with possible and probable causes in banking and lead to greater customer satisfaction and operational efficiency.

One of the most significant areas with promising applications is in handling complex, ambiguous queries with high accuracy. This opens up a possibility for banks to automate a bigger proportion of customer interactions without sacrificing quality of service. The better scores of customers and the lower rates of escalation further emphasize the potential of AI in redefining bank's customer service.

Still, AI-based chatbot systems in banking pose a lot of challenges. Questions of transparency about designing models, the issue of accuracy, and ethical considerations require extreme caution. The potential avenues for future research would include developing more interpretable AI models, developing robust monitoring and retraining protocols, and ways to mitigate the biases of AI systems in their intended applications.

Since our study was text-based in nature, further work can be conducted in the integration of AI chatbots with voice recognition technology, incorporating new avenues for customer interaction. Further research examining the long-term effects of AI chatbots on customer loyalty and performance of banks might help the industry gain better insights.

Conclusion Based on what has been discovered, generative AI could thus be the perfect answer to creating an optimal solution for retail banking customer support chatbots. Reduced resolution times, along with improvements in customer satisfaction and operational efficiency, are therefore poised to dramatically alter the future of banking customer service-and it will be key for banks to monitor the development of this technology and apply careful judgment to how AI can best be integrated into their customer service strategies.

References

[1] Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. Machine Learning with Applications, 2, 100006.

[2] Atay, E., Türkyılmaz, İ., & Ünver, M. (2020). Chatbot implementation in banking sector: An empirical study on customer satisfaction. International Journal of Advanced Research in Engineering and Technology, 11(4), 278-289.

[3] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610-623).

[4] Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

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

[6] Chaves, A. P., & Gerosa, M. A. (2021). How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. International Journal of Human–Computer Interaction, 37(8), 729-758.

[7] Cui, L., Huang, S., Wei, F., Tan, C., Duan, C., & Zhou, M. (2017). SuperAgent: A customer service chatbot for e-commerce websites. In Proceedings of ACL 2017, System Demonstrations (pp. 97-102).

[8] Dale, R. (2016). The return of the chatbots. Natural Language Engineering, 22(5), 811-817.

[9] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter

of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 4171-4186).

[10] Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of HCI. Interactions, 24(4), 38-42.

[11] Gnewuch, U., Morana, S., & Maedche, A. (2017). Towards designing cooperative and social conversational agents for customer service. In Proceedings of the International Conference on Information Systems (ICIS).

[12] Huang, J., Zhou, M., & Yang, D. (2007). Extracting chatbot knowledge from online discussion forums. In Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI) (pp. 423-428).

[13] Hussain, S., Sianaki, O. A., & Ababneh, N. (2019). A survey on conversational agents/chatbots classification and design techniques. In Web, Artificial Intelligence and Network Applications (pp. 946-956). Springer, Cham.

[14] Hutto, C. J., & Gilbert, E. (2020). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 8, No. 1).

[15] Jain, M., Kumar, P., Kota, R., & Patel, S. N. (2018). Evaluating and informing the design of chatbots. In Proceedings of the 2018 Designing Interactive Systems Conference (pp. 895-906).

[16] Jurafsky, D., & Martin, J. H. (2020). Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition. Pearson Education.

[17] Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M. E., ... & Staab, S. (2020). Bias in data-driven artificial intelligence systems—An introductory survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3), e1356.

[18] Przegalinska, A., Ciechanowski, L., Stroz, A., Gloor, P., & Mazurek, G. (2019). In bot we trust: A new methodology of chatbot performance measures. Business Horizons, 62(6), 785-797.

[19] Reiter, E. (2018). A structured review of the validity of BLEU. Computational Linguistics, 44(3), 393-401.

[20] Roller, S., Dinan, E., Goyal, N., Ju, D., Williamson, M., Liu, Y., ... & Weston, J. (2021). Recipes for building an open-domain chatbot. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume (pp. 300-325).

[21] Toxtli, C., Monroy-Hernández, A., & Cranshaw, J. (2018). Understanding chatbot-mediated task management. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (pp. 1-6).

[22] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems (pp. 5998-6008).

[23] Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A new chatbot for customer service on social media. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 3506-3510). ACM.

[24] Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A new chatbot for customer service on social media. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 3506-3510).

[25] Zumstein, D., & Hundertmark, S. (2017). Chatbots: An interactive technology for personalized communication and transaction. International Journal of WWW/Internet, 15(1), 96-109.