Care-by-Design: Enhancing Customer Experience through Empathetic AI Chatbots

Kadhim Hayawi
AICEO
Arjaan Office Tower 302, AL Sufouh Complex, Tecom
Dubai, UAE
kadhim.hayawi@ceoai.ai

Sakib Shahriar
AICEO
Arjaan Office Tower 302, AL Sufouh Complex, Tecom
Dubai, UAE
shahrias@uoguelph.ca

***Abstract*— The reliance on AI-driven customer service has improved efficiency but often lacks the empathy and personalization that foster deeper customer relationships. Traditional customer service models are limited by human agents' capacity to remember past interactions and adapt to individual customer needs in real-time. This paper introduces a Care-by-Design (CbD) framework that embeds empathy, contextual learning, and personalization into the training and deploying large language models (LLMs) for customer service. The proposed framework emphasizes building meaningful customer relationships by addressing emotional and contextual cues in customer interactions. To validate this approach, we conducted a case study comparing a baseline LLM with a CbD-enhanced version fine-tuned on an empathy-focused dataset. Our results demonstrated improved customer satisfaction, query understanding, and emotional responsiveness in the CbD model.**

Keywords—Emotion AI, Large Language Models, Customer Service Automation, Contextual Learning, Chatbots

I. Introduction

Empathy and care in human relationships have long been recognized as fundamental to building trust, fostering loyalty, and ensuring mutual understanding [1]. Therefore, customer service has traditionally focused on empathy and care to meet customer satisfaction [2]. Before the rise of artificial intelligence (AI) and modern technology, businesses relied on human agents to provide personal, attentive service tailored to individual customer needs. Call centers and in-person interactions were structured around the idea that a customer's concerns should be heard, understood, and addressed with care [3]. However, human limitations, such as the inability to remember past interactions and the finite emotional and cognitive bandwidth required to provide continuous care, have posed challenges in delivering personalized and empathetic service [4]. Memory fades, and no human agent can handle vast interactions without fatigue or emotional strain, leading to inconsistencies in customer care.

This human-centric model faced further limitations as customer interactions moved online and scaled globally. During the COVID-19 pandemic, for example, the limitations of human-driven customer service became even more pronounced. A study conducted on call center agents in the Philippines found that the heightened demands and stress during the pandemic led to significant job burnout, which in turn negatively impacted job satisfaction [5]. This burnout affected the agents' well-being and compromised the quality of care they could provide. The emotional toll and high-pressure environment made it difficult for human agents to consistently provide empathetic responses. While chatbots and automated systems have introduced efficiency, they have often fallen short of replicating the empathy central to meaningful customer care [6]. In such scenarios, AI systems capable of delivering care-focused, empathetic responses would alleviate some of the burdens from human agents and ensure a more consistent and emotionally intelligent customer service experience.

The evolution from traditional machine learning and AI, such as predictive algorithms [7], to generative AI [8] has opened new possibilities for designing a care-centric service model. Traditional AI systems excelled at specific tasks like predicting customer behavior, automating routine inquiries, and recommending products based on predefined patterns ([9], [10]). However, these models lacked the flexibility and depth to engage in nuanced, empathetic conversations. Predictive algorithms could anticipate what a customer might want but struggled to address the emotional context behind those needs. With the advent of generative AI and large language models (LLMs) [11], we can move beyond task-based automation toward more human-like interactions. LLMs can understand and generate language that responds not only to the content of customer queries but also to their emotional and contextual undertones. This capacity to dynamically generate empathetic responses allows AI to replicate and enhance the care traditionally offered by human agents, leading to a care-centric service experience.

Building on the unique capabilities of LLMs, we advocate for a *care-by-design* approach in training AI systems, where empathy and care are embedded into every aspect of the technology design. A conceptual framework is introduced in this work that outlines how we can integrate care systematically into LLMs using methods such as fine-tuning and prompt engineering. To demonstrate the practical effectiveness of this approach, we also provide a case study that showcases how these methods can transform AI-driven customer service into a care-centric experience.

II. Problem Statement

A. *Limitations of Human-Centric Customer Service*

Traditional customer service models, especially those centered around human agents, often prioritize logical and rule-based responses over empathetic engagement. As depicted in Figure 1, when a customer makes a unique request asking for a purple debit card, the typical response from a customer service representative is a straightforward rejection: "Sorry, we don't offer purple debit cards." This transactional approach addresses the immediate query but lacks an understanding of the emotional needs or the context behind the request. Such interactions, focused purely on fulfilling or denying requests, miss opportunities to engage customers in a more meaningful, care-centric way. Without flexibility or creative problem-solving, these responses can feel cold and dismissive, leaving the customer dissatisfied. The absence of empathy and out-of-the-box thinking in these interactions reveals a key limitation of human-centered customer service, particularly when agents face burnout and stress, or are restricted by rigid policies.

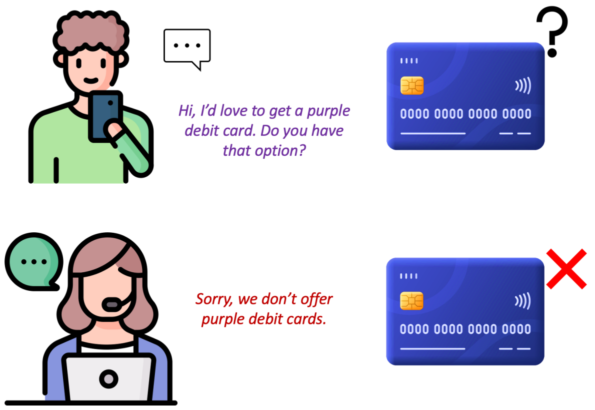

Figure 1 Illustrative example: the purple debit card scenario

B. *Towards a Empathatic Approach*

LLMs offer a promising alternative to traditional customer service by enabling more care-driven interactions. As shown in Figure 2, these models can detect emotions by analyzing the tone and content of a user request. For instance, if a customer asks for a purple debit card, the model can infer the emotional significance behind the color choice. Unlike human agents who may forget past conversations, LLMs can retain context from previous interactions, providing a deeper understanding of a customer's preferences, as seen in the model recalling a customer's love for unique items. Moreover, the model can offer creative alternatives (like suggesting a custom cover) while showing empathy; the customer feels heard and valued even if the exact request is not fulfilled.

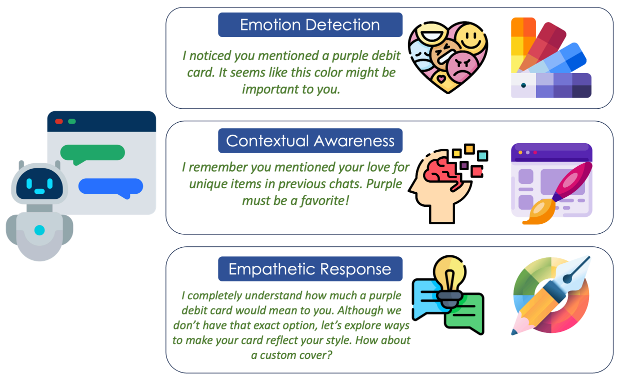

Figure 2 Generative AI responding to the purple debit card scenario

The implications of this care-centric approach go beyond simply resolving customer queries; it fosters a deeper connection between the customer and the brand. As illustrated in Figure 3, when customers are involved in value co-creation [12], such as choosing a custom design for their debit card, they feel a greater sense of ownership and partnership with the company. This personalized interaction makes the customer feel valued as part of the company's extended family. Even if their initial request for a specific product cannot be met exactly, offering alternative solutions that align with their preferences can significantly enhance their satisfaction. Organizations can cultivate *"brand love"* using this approach, where customers develop a stronger emotional bond and loyalty to the company. Consequently, this leads to higher customer satisfaction, retention, and positive word-of-mouth.

Figure 3 Value co-creation and customer satisfaction through imaginative solutions

III. Care-by-Design: A Conceptual Framework

The proposed *Care-by-Design (CbD)* framework integrates empathy and personalized care into designing and deploying LLMs for customer service. The framework emphasizes the importance of embedding care into every stage of technology development, from the foundational training of LLMs to real-time interaction and ongoing refinement. In traditional AI models, efficiency and task fulfillment were prioritized, leaving emotional engagement and relationship-building secondary concerns. However, with the rise of generative AI like LLMs, there is an opportunity to transform customer service by incorporating emotional intelligence, contextual awareness, and dynamic personalization. The CbD framework is built around eight key components to enhance the model's ability to provide empathetic and personalized customer interactions.

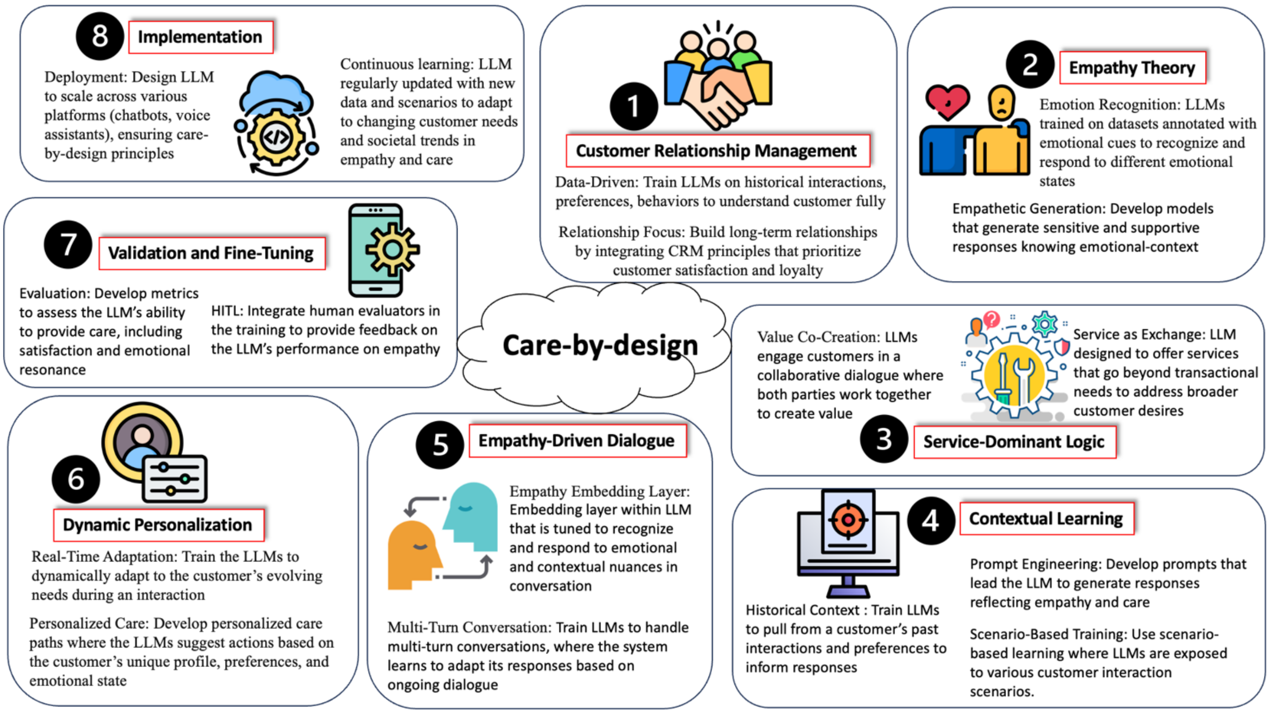

Figure 4 Care-by-Design conceptual framework for enhancing empathy and personalizes care in customer service

A. Customer Relationship Management

Customer Relationship Management (CRM) emphasizes building and maintaining long-term customer relationships. Traditionally, CRM relied on human agents to remember customer preferences, handle multiple interactions, and use those experiences to foster a sense of connection and loyalty [13]. However, humans have a limited capacity to recall details from past conversations or recognize patterns in customer behavior over time. Moreover, unconscious biases may affect how agents interact with customers from different backgrounds, leading to inconsistent service quality. For example, a study demonstrated that interactions between service employees and customers from various ethnic backgrounds (interethnic encounters) often result in less positive behavior from majority-group customers than interactions with same-ethnicity service providers [14]. The integration of CRM principles into LLMs can overcome these limitations. LLMs can be trained to retain and recall vast amounts of customer data like past interactions, preferences, purchasing history, and even emotional tones from previous conversations [15]. LLMs can analyze this data in real-time to derive patterns and offer personalized responses tailored to each customer's unique needs. Additionally, by removing human biases, AI-driven CRM can ensure that all customers, regardless of their background, receive consistent, fair, and empathetic service.

B. Empathy Theory

Human agents often struggle to pick up on subtle emotional cues or respond empathetically in high-stress situations. Factors like fatigue or workload may affect their ability to engage with user feelings [16]. Empathy is essential for building trust and rapport, but it can be challenging for humans to consistently demonstrate empathy, especially in emotionally charged or repetitive interactions. Empathy Theory focuses on LLMs recognizing and responding to emotions in variable scenarios, i.e., cognitive empathy [17]. LLMs can be trained to detect emotions from the language used by customers. For instance, if a customer expresses frustration by saying, "I'm really upset that my order hasn't arrived," a care-centric LLM can recognize the emotional context and respond with understanding. Rather than a robotic or generic response, the LLM might say, "I completely understand how frustrating this delay must be. Let me see how I can resolve this for you quickly." This type of empathetic response isn't just about solving the problem; it's about making the customer feel heard and valued.

C. Service-Dominant Logic

Traditionally, customer service has often been reactive; customers reach out with a problem, and agents offer solutions within the company's predefined boundaries. However, this approach limits the potential for building meaningful customer relationships and addressing broader customer needs. Service-dominant logic (SDL) shifts the focus from delivering products or services as mere transactions to co-creating value with customers through interactive exchanges [18]. SDL emphasizes that services are about fulfilling requests and co-creating value in partnership with the customer. For example, if a customer wants an unavailable product or asks for a feature the company does not currently offer (like requesting a purple debit card!), the LLM can engage the customer in a dialogue to understand the deeper motivations behind the request. Perhaps the customer is looking for more personalization in their banking experience. The service-as-exchange approach makes customer interactions from transactional problem-solving to collaborative problem-solving. We can exploit an LLM's contextual understanding and dynamic response generation capabilities to provide valuable customer services [19].

D. Contextual Learning

In traditional customer service, human agents may struggle to remember details from previous conversations or understand the specific preferences of returning customers. Even with notes or CRM systems, it's difficult for human agents to access and synthesize vast amounts of customer data in real time. This often results in disconnected experiences, where customers feel like they have to re-explain their needs every time they interact with a company [20]. Interpretable knowledge distillation can enhance contextual learning in LLMs by teaching smaller models to learn adaptive strategies from larger models without directly manipulating their parameters [21]. This method creates a library of customer interaction scenarios and optimal strategies, enabling LLMs to remember and apply past customer preferences in real-time interactions. For example, if a customer has previously expressed a preference for eco-friendly products, the LLM can take this into account during a new interaction and offer suggestions, even if the customer hasn't explicitly asked for them. Additionally, by utilizing prompt engineering, the LLM can be guided to generate empathetic responses [22]. Prompts such as "*How can we offer care and support based on the customer's previous interaction?*" ensure that the LLM provides responses that reflect the customer's emotional state. Moreover, scenario-based training allows the LLM to learn from various customer interaction scenarios, preparing it to handle different inquiries and emotional contexts [23].

E. Empathy-Driven Dialogue

Even well-trained human agents can struggle to consistently maintain empathy when dealing with high volumes of inquiries or challenging situations. Fatigue, personal bias, or time pressure can lead to interactions that feel transactional rather than understanding. Empathy-driven dialogue in LLMs addresses this challenge by integrating an empathy embedding layer into the model [24]. For instance, if a customer expresses frustration over a delayed order, the LLM can detect not only the content of the complaint but also the underlying emotion. The responses can be further refined with a neural language model postprocessing [25]. In addition, LLMs can be trained for multi-turn conversations, where they adapt their responses based on the ongoing dialogue [26], leading the LLM to address the query, follow up, or check in on unresolved issues to offer additional assistance.

F. Dynamic Personalization

Human agents often have limited capacity to personalize each interaction on an ongoing basis. While they may have access to customer profiles or past interactions, agents can struggle to adapt to the changing emotions, preferences, or instantaneous user needs. Dynamic personalization allows LLMs to tailor their responses in real time based on the customer's current behavior, past interactions, and expressed preferences [27]. This adaptability is attained by real-time data analysis and the LLM's ability to adjust its responses as the conversation progresses. Personalized care paths allow the LLM to go beyond merely addressing the current query to offering thoughtful suggestions that anticipate the customer's future needs.

G. Validation and Fine-tuning

Human agents receive regular training and performance feedback to improve their customer interactions. Similarly, LLMs should undergo regular validation and fine-tuning based on qualitative and quantitative measures. Key metrics include customer satisfaction, emotional resonance, and the LLM's ability to provide thoughtful and empathetic solutions. Through *human-in-the-loop (HITL)* integration, human evaluators can be involved in the training and validation process, offering feedback on the LLM's performance in real-world interactions [28]. These evaluators can assess how well the LLM recognizes and responds to emotional cues, provides personalized care, and resolves customer concerns effectively. Fine-tuning also allows the LLM to adapt to new customer needs and societal trends. As customer expectations evolve, the LLM can be updated with new datasets, scenarios, and customer interaction patterns to ensure it stays relevant and empathetic in its responses. This iterative process helps prevent the model from becoming stagnant and continues to offer a care-centric experience.

H. Implementation

Implementation involves designing the LLM to function seamlessly across different platforms, whether chatbots on a website, voice assistants, or messaging services. The key is maintaining empathy and care, regardless of how or where customers engage with the service. A crucial aspect of implementation is continuous learning. Customer preferences, societal trends, and expectations are constantly evolving. Thus, updating the model periodically with new data, customer interactions, and emotional scenarios is needed to ensure the model stays relevant and empathetic. For instance, during periods of social or economic change, the LLM may need to adjust its tone or suggestions to be more sensitive to customers' current realities. Finally, care-by-design principles must be maintained throughout the LLM's lifecycle. The deployment process should include regular audits [29] so that care values are not compromised when adding new features.

IV. CASE STUDY

The primary objective of this case study is to assess the effectiveness of implementing Care-by-Design (CbD) principles in enhancing customer satisfaction through AI-driven customer service interactions. Specifically, this study aims to evaluate the impact of fine-tuning an LLM on an empathy-focused dataset and using prompt engineering to improve the model's ability to detect emotions, provide personalized care, and foster deeper customer relationships. The case study compares a base model (Model A) and a model customized with CbD principles (Model B), analyzing the effects of these modifications on customer satisfaction.

A. Methods

For this study, we fine-tuned GPT-4 [30] using the BlendedSkillTalk dataset [31] of 7,000 customer service conversations, carefully curated to exhibit multiple conversation modes: empathy, personality, and knowledge. In contrast, the baseline model was the standard GPT-4 model without additional fine-tuning or modifications. In addition to fine-tuning, we employed prompt engineering in Model B to enhance its responses with a more empathetic and personalized approach. The types of prompts included phrases like "Try to understand user feelings and emotions" and "Be a caring friend."

B. Study Design

To assess the effectiveness of the CbD approach, we recruited 20 participants to interact with both models. Participants were divided into two groups:

- Group 1 (10 participants): Interacted with base GPT-4 model (baseline).
- Group 2 (10 participants): Interacted with GPT-4 fine-tuned and prompt-engineered with CbD principles (CbD).

Each participant was given a link to a chatbot interface where they could interact with the models, asking service-related questions just as they would in a real customer service scenario. Participants were encouraged to ask at least 3-5 questions to evaluate the LLM's responses and empathetic engagement. Participants were guided to ask questions across diverse scenarios spanning requests, complaints, feedback, and inquiries. Sample questions included "Can I change the delivery address for my order?" and "Your website is hard to navigate. It takes too long to find what I need." However, the respondents could also ask other questions or comments that reflected typical customer service concerns.

C. *Results*

After interacting with the chatbot, participants filled out a survey that evaluated their overall experience, focusing on:

1. Response relevance and clarity
2. Perceived empathy and emotional intelligence
3. Satisfaction with the chatbot's handling of their inquiry
4. Likelihood of using the chatbot again for future customer service needs

The objective was to examine whether the CbD model outperformed the base model in fostering customer satisfaction. Participants also provided valuable insights into the perceived empathy and care in the chatbot's responses through optional comments. We summarize the results from the case study in Table 1 and Figure 5.

Table 1 Comparison in user satisfaction between baseline and CbD models

	Baseline	CbD
Overall Experience	2.7	**3.3**
Overall Satisfaction	2.6	**3.0**
Understand Query	2.2	**2.9**
Understand Emotion	3.1	2.6
Acknowledge Feelings	2.2	**2.6**
Address Needs	2.7	**3.2**
Demonstrate Care	2.9	**3.3**

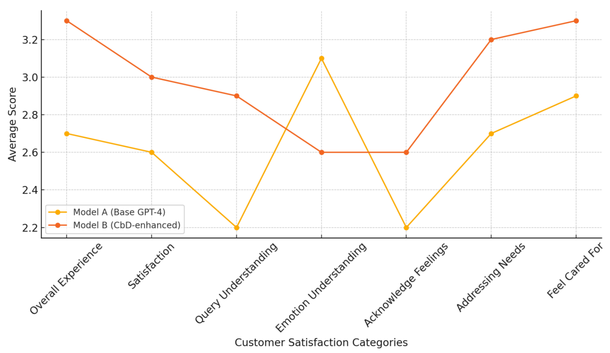

Figure 5 Comparison of baseline and CbD models on user satisfaction

The analysis of the results shows a noticeable distinction between the performance of the baseline and the CbD model across various customer satisfaction metrics. The CbD model outperformed the baseline in areas such as overall experience and satisfaction, suggesting that fine-tuning an empathy dataset and prompt engineering improved the quality of customer interactions. User comments also support this conclusion. For instance, one participant noted, "The chatbot shows empathy in its responses" when interacting with the CbD model. Another participant mentioned a "good experience fast response from the bot," reinforcing the improved responsiveness and care that CbD offers. Conversely, the baseline model received feedback like "The chatbot gives generic answers," reflecting a less personalized and more robotic experience. This response aligns with the lower scores Model A received in categories like Emotion Understanding and Addressing Needs, where it fell behind the more empathetic and adaptable Model B. These results demonstrate the effectiveness of CbD principles in creating a more engaging, caring, and satisfactory experience.

V. Discussion

Despite the promise of CbD approach in providing empathetic service, several factors should be considered. In this section, we briefly discuss several considerations for CbD moving forward.

A. *Automated Systems and Human Agents in Customer Service*

In terms of their capabilities of emotionally engaging customers, LLM-based approaches have the potential to provide more consistent emotional recognition and empathy responses. Automated systems have been proven to reduce human biases and variability [32]; however, they may lack contextual depth when it comes to human understanding of intricate emotional scenarios [33]. Research has also demonstrated that service robots can illicit positive emotions like excitement but they struggle with complex emotions in malfunctioning scenarios [34].

An essential consideration in customer service is responsiveness. To this end, automated systems like AI-based chatbots are considered highly scalable with their ability to work with massive data volumes and yet providing real-time responses [33]. In contrast, human agents may offer slower responses due to fatigue or other limitations although they are more capable in adapting to unforeseen circumstances. In this context, human empathy is more flexible and nuanced in effectively adjusting based on a customer's variable emotional state [35].

When it comes to providing accurate responses, general AI systems have mixed successes in different criteria, like extracting latent topics and emotion detection [36]. However, the advent of LLM has offered promise in this context as they are considered more accurate in judgments and evaluations. Human agents on the other hand may outperform AI systems, particularly during scenarios that need complex decision-making due to the presence of ambiguity or cases when emotions highly dictate a customer's decision-making [37].

Studies show that while AI is effective in more mechanical or thinking-based tasks, human intelligence is often preferred for "feeling-based" service interactions [33]. The ability of human agents to connect emotionally, particularly in high-stakes or sensitive situations, still provides a competitive edge over AI-based systems [38]. While AI-driven systems provide unmatched scalability [35], delivering personalized responses at scale remains a barrier. Human agents, though limited in scalability, excel in customized care by adapting their responses based on the customer's individual needs and emotional state. As the technological shift continues with the arrival of generative AI and augmented reality, automated systems are quickly closing the service gap [39].

B. Ethical Considerations

Implementing automated and AI-based services using CbD should consider ethical dilemmas. It is well known that many AI models inherit biases from training data [40]. For example, AI-driven emotion recognition could misinterpret emotions based on cultural or linguistic differences. This could lead to unequal treatment of customers based on factors like race, gender, and socio-economic status. AI models require extensive data to improve performance. This further raises concerns over how personal customer data is collected, stored, and processed. For example, AI models trained on personal data could be prone to attacks like model inference and poisoning attacks by adversaries seeking to obtain personal information from models [29]. Strong privacy policies and compliance with data protection regulations, such as GDPR, are necessary [29]. Therefore, to manage the ethical dilemmas, there should be human oversight in developing AI in automation, which would ensure appropriate responses in complex and emotionally charged situations.

VI. Conclusion

This research demonstrates the effectiveness of implementing Care-by-Design (CbD) principles, such as fine-tuning the model on empathy datasets and using prompt engineering, to enhance the customer experience in AI-driven chatbots. Through a case study, we proved that incorporating CbD principles showed improvements in overall user satisfaction, understanding of customer needs, and emotional responsiveness. User feedback, which emphasized the chatbot's empathetic and personalized responses, further validates the benefits of integrating care-focused design into AI systems. Future studies should explore the full implementation of the CbD framework by incorporating all eight components, including dynamic personalization and continuous learning. Multimodal AI systems, which can process and integrate multiple data types, offer a promising solution for delivering more comprehensive and personalized customer care [41]. Additionally, large-scale user testing with a more diverse participant pool would provide further insights into the scalability and robustness of CbD-enhanced models in real-world applications.

References

[1] N. Sharma and P. G. Patterson, “The impact of communication effectiveness and service quality on relationship commitment in consumer, professional services,” *J. Serv. Mark.*, vol. 13, no. 2, pp. 151–170, Jan. 1999, doi: 10.1108/08876049910266059.

[2] D. J. Shemwell, U. Yavas, and Z. Bilgin, “Customer-service provider relationships: an empirical test of a model of service quality, satisfaction and relationship-oriented outcomes,” *Int. J. Serv. Ind. Manag.*, vol. 9, no. 2, pp. 155–168, Jan. 1998, doi: 10.1108/09564239810210505.

[3] R. Feinberg, K. de Ruyter, and L. Bennington, *Cases in Call Center Management: Great Ideas (th)at Work*. Purdue University Press, 2005.

[4] F. Yakob, *Paid Attention: Innovative Advertising for a Digital World*. Kogan Page Publishers, 2021.

[5] A. Santiago, C. Barandino, A. Bil, R. Curam, K. Torrero, and J. Tus, “Call Center Agents’ Job Burnout and Its’ Influence on their Job Satisfaction During the COVID-19 Pandemic in the Philippines,” *Int. J. Adv. Res. Innov. Ideas Educ.*, vol. 7, p. 2021, Jun. 2021, doi: 10.6084/m9.figshare.14810592.v1.

[6] C. Crolic, F. Thomaz, R. Hadi, and A. T. Stephen, “Blame the Bot: Anthropomorphism and Anger in Customer–Chatbot Interactions,” *J. Mark.*, vol. 86, no. 1, pp. 132–148, Jan. 2022, doi: 10.1177/00222429211045687.

[7] S. Shahriar, A. R. Al-Ali, A. H. Osman, S. Dhou, and M. Nijim, “Machine Learning Approaches for EV Charging Behavior: A Review,” *IEEE Access*, vol. 8, pp. 168980–168993, 2020, doi: 10.1109/ACCESS.2020.3023388.

[8] N. R. Mannuru *et al.*, “Artificial intelligence in developing countries: The impact of generative artificial intelligence (AI) technologies for development,” *Inf. Dev.*, p. 02666669231200628, Sep. 2023, doi: 10.1177/02666669231200628.

[9] I. A. Zualkernan, M. Pasquier, S. Shahriar, M. Towheed, and S. Sujith, “Using BLE beacons and Machine Learning for Personalized Customer Experience in Smart Cafés,” in *2020 International Conference on Electronics, Information, and Communication (ICEIC)*, Jan. 2020, pp. 1–6. doi: 10.1109/ICEIC49074.2020.9051187.

[10] S. Shahriar and A. R. Al-Ali, “Impacts of COVID-19 on Electric Vehicle Charging Behavior: Data Analytics, Visualization, and Clustering,” *Appl. Syst. Innov.*, vol. 5, no. 1, Art. no. 1, Feb. 2022, doi: 10.3390/asi5010012.

[11] K. Hayawi, S. Shahriar, H. Alashwal, and M. A. Serhani, “Generative AI and large language models: A new frontier in reverse vaccinology,” *Inform. Med. Unlocked*, vol. 48, p. 101533, Jan. 2024, doi: 10.1016/j.imu.2024.101533.

[12] Z. Piligrimiene, A. Dovaliene, and R. Virvilaite, “Consumer Engagement in Value Co-Creation: what Kind of Value it creates for Company?,” *Eng. Econ.*, vol. 26, no. 4, Art. no. 4, Oct. 2015, doi: 10.5755/j01.ee.26.4.12502.

[13] J. Dyché, *The CRM Handbook: A Business Guide to Customer Relationship Management*. Addison-Wesley Professional, 2002.

[14] M. Linzmajer, S. Brach, G. Walsh, and T. Wagner, “Customer Ethnic Bias in Service Encounters,” *J. Serv. Res.*, vol. 23, no. 2, pp. 194–210, May 2020, doi: 10.1177/1094670519878883.

[15] Y. Hou, H. Tamoto, and H. Miyashita, “‘My agent understands me better’: Integrating Dynamic Human-like Memory Recall and Consolidation in LLM-Based Agents,” in *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems*, in CHI EA ’24. New York, NY, USA: Association for Computing Machinery, May 2024, pp. 1–7. doi: 10.1145/3613905.3650839.

[16] C. Dormann and D. Zapf, “Customer-Related Social Stressors and Burnout,” *J. Occup. Health Psychol.*, vol. 9, pp. 61–82, Jan. 2004, doi: 10.1037/1076-8998.9.1.61.

[17] V. Sorin *et al.*, “Large Language Models (LLMs) and Empathy – A Systematic Review,” Aug. 07, 2023, *medRxiv*. doi: 10.1101/2023.08.07.23293769.

[18] A. Tommasetti, O. Troisi, and M. Vesci, “Measuring customer value co-creation behavior,” *J. Serv. Theory Pract.*, vol. 27, no. 5, pp. 930–950, Jan. 2017, doi: 10.1108/JSTP-10-2015-0215.

[19] S. Kolasani, “Optimizing Natural Language Processing, Large Language Models (LLMs) for Efficient Customer Service, and hyper-personalization to enable sustainable growth and revenue,” *Trans. Latest Trends Artif. Intell.*, vol. 4, no. 4, Jul. 2023, Accessed: Sep. 10, 2024. [Online]. Available: https://ijsdcs.com/index.php/TLAI/article/view/476

[20] M. Lajante and M. Del Prete, “Technology-infused Organizational Frontlines: When (Not) to Use Chatbots in Retailing to Promote Customer Engagement,” in *Retail Futures*, E. Pantano, Ed., Emerald Publishing Limited, 2020, pp. 71–84. doi: 10.1108/978-1-83867-663-620201011.

[21] T. Wang, K. Sudhir, and D. Hong, “Using Advanced LLMs to Enhance Smaller LLMs: An Interpretable Knowledge Distillation Approach,” Aug. 13, 2024, *arXiv*: arXiv:2408.07238. doi: 10.48550/arXiv.2408.07238.

[22] M. Thanasi-Boçe and J. Hoxha, “From ideas to ventures: building entrepreneurship knowledge with LLM, prompt engineering, and conversational agents,” *Educ. Inf. Technol.*, Jun. 2024, doi: 10.1007/s10639-024-12775-z.

[23] T. Yamazaki, K. Yoshikawa, T. Kawamoto, T. Mizumoto, M. Ohagi, and T. Sato, “Building a hospitable and reliable dialogue system for android robots: a scenario-based approach with large language models,” *Adv. Robot.*, vol. 37, no. 21, pp. 1364–1381, Nov. 2023, doi: 10.1080/01691864.2023.2244554.

[24] A. S. Raamkumar and Y. Yang, “Empathetic Conversational Systems: A Review of Current Advances, Gaps, and Opportunities,” *IEEE Trans. Affect. Comput.*, vol. 14, no. 4, pp. 2722–2739, Oct. 2023, doi: 10.1109/TAFFC.2022.3226693.

[25] A. Vázquez, A. López Zorrilla, J. M. Olaso, and M. I. Torres, "Dialogue Management and Language Generation for a Robust Conversational Virtual Coach: Validation and User Study," *Sensors*, vol. 23, no. 3, Art. no. 3, Jan. 2023, doi: 10.3390/s23031423.

[26] Z. Teng, Y. Song, X. Ye, and Y. Ouyang, "Fine-Tuning LLMs for Multi-Turn Dialogues: Optimizing Cross-Entropy Loss with KL Divergence for All Rounds of Responses," in *Proceedings of the 2024 16th International Conference on Machine Learning and Computing*, in ICMLC '24. New York, NY, USA: Association for Computing Machinery, Jun. 2024, pp. 128–133. doi: 10.1145/3651671.3651702.

[27] S. Ghodratnama and M. Zakershahrak, "Adapting LLMs for Efficient, Personalized Information Retrieval: Methods and Implications," in *Service-Oriented Computing – ICSOC 2023 Workshops*, F. Monti, P. Plebani, N. Moha, H. Paik, J. Barzen, G. Ramachandran, D. Bianchini, D. A. Tamburri, and M. Mecella, Eds., Singapore: Springer Nature, 2024, pp. 17–26. doi: 10.1007/978-981-97-0989-2_2.

[28] C. O. Retzlaff *et al.*, "Human-in-the-Loop Reinforcement Learning: A Survey and Position on Requirements, Challenges, and Opportunities," *J. Artif. Intell. Res.*, vol. 79, pp. 359–415, Jan. 2024, doi: 10.1613/jair.1.15348.

[29] S. Shahriar, S. Allana, S. M. Hazratifard, and R. Dara, "A Survey of Privacy Risks and Mitigation Strategies in the Artificial Intelligence Life Cycle," *IEEE Access*, vol. 11, pp. 61829–61854, 2023, doi: 10.1109/ACCESS.2023.3287195.

[30] OpenAI *et al.*, "GPT-4 Technical Report," Mar. 04, 2024, *arXiv*: arXiv:2303.08774. doi: 10.48550/arXiv.2303.08774.

[31] E. M. Smith, M. Williamson, K. Shuster, J. Weston, and Y.-L. Boureau, "Can You Put it All Together: Evaluating Conversational Agents' Ability to Blend Skills," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault, Eds., Online: Association for Computational Linguistics, Jul. 2020, pp. 2021–2030. doi: 10.18653/v1/2020.acl-main.183.

[32] L. Xiao and V. Kumar, "Robotics for Customer Service: A Useful Complement or an Ultimate Substitute?," *J. Serv. Res.*, vol. 24, no. 1, pp. 9–29, Feb. 2021, doi: 10.1177/1094670519878881.

[33] M.-H. Huang and R. T. Rust, "Artificial Intelligence in Service," *J. Serv. Res.*, vol. 21, no. 2, pp. 155–172, May 2018, doi: 10.1177/1094670517752459.

[34] S. H.-W. Chuah and J. Yu, "The future of service: The power of emotion in human-robot interaction," *J. Retail. Consum. Serv.*, vol. 61, p. 102551, Jul. 2021, doi: 10.1016/j.jretconser.2021.102551.

[35] S. E. Sampson, "A Strategic Framework for Task Automation in Professional Services," *J. Serv. Res.*, vol. 24, no. 1, pp. 122–140, Feb. 2021, doi: 10.1177/1094670520940407.

[36] R. Filieri, Z. Lin, Y. Li, X. Lu, and X. Yang, "Customer Emotions in Service Robot Encounters: A Hybrid Machine-Human Intelligence Approach," *J. Serv. Res.*, vol. 25, no. 4, pp. 614–629, Nov. 2022, doi: 10.1177/10946705221103937.

[37] N. Warren, S. Hanson, and H. Yuan, "Feeling Manipulated: How Tip Request Sequence Impacts Customers and Service Providers?," *J. Serv. Res.*, vol. 24, no. 1, pp. 66–83, Feb. 2021, doi: 10.1177/1094670519900553.

[38] F. McLeay, V. S. Osburg, V. Yoganathan, and A. Patterson, "Replaced by a Robot: Service Implications in the Age of the Machine," *J. Serv. Res.*, vol. 24, no. 1, pp. 104–121, Feb. 2021, doi: 10.1177/1094670520933354.

[39] J. Heller, M. Chylinski, K. de Ruyter, D. I. Keeling, T. Hilken, and D. Mahr, "Tangible Service Automation: Decomposing the Technology-Enabled Engagement Process (TEEP) for Augmented Reality," *J. Serv. Res.*, vol. 24, no. 1, pp. 84–103, Feb. 2021, doi: 10.1177/1094670520933692.

[40] N. Norori, Q. Hu, F. M. Aellen, F. D. Faraci, and A. Tzovara, "Addressing bias in big data and AI for health care: A call for open science," *Patterns*, vol. 2, no. 10, Oct. 2021, doi: 10.1016/j.patter.2021.100347.

[41] S. Shahriar *et al.*, "Putting GPT-4o to the Sword: A Comprehensive Evaluation of Language, Vision, Speech, and Multimodal Proficiency," *Appl. Sci.*, vol. 14, no. 17, Art. no. 17, Jan. 2024, doi: 10.3390/app14177782.