Chatbot Systems: Measuring Effectiveness, Challenges, and Future Directions Across Industries

Edmira Xhaferra

Amarildo Rista

Lindita Mukli

Fatmira Prodani

Frida Gjermeni

Faculty of Information Technology,
Aleksander Moisiu University of Durres,
Durres, Albania

Received: 24 May 2025 / Accepted: 24 July 2025 / Published: 01 August 2025

Doi: 10.56345/ijrdv12n2013

Abstract

Chatbot systems have become a transformative technology reshaping customer interaction across multiple industries. This study provides a comprehensive analysis of chatbot implementation, focusing on their effectiveness in improving operational efficiency, enhancing user satisfaction, and reducing organizational costs. Drawing upon secondary data and three case studies Erica in finance, Woebot in healthcare, and Sephora Virtual Artist in retail the study evaluates quantitative and qualitative outcomes of chatbot adoption. Results indicate significant improvements, including reductions in service workload, increased customer satisfaction, and higher engagement metrics. Despite these gains, challenges such as data privacy compliance, bias in AI models, and technical integration complexities persist. The study emphasizes the potential of emerging technologies like generative AI and predictive analytics to address these limitations and enhance chatbot capabilities further. It concludes by recommending strategic investments in advanced NLP, backend integration, and cultural adaptability for businesses deploying chatbot systems. Future research should explore longitudinal impacts and ethical frameworks to sustain long-term effectiveness.

Keywords: *chatbot systems, artificial intelligence, natural language processing, customer satisfaction, data privacy, and generative AI*

1. Introduction

Chatbot systems represent a transformative technology within the digital realm, shifting how companies interact with users. Generally speaking, chatbots are classified into two: rule-based and AI-driven. Rule-based chatbots work off of pre-established scripts and decision trees, rendering them inept to deal with any complex conversations. The AI-driven counterpart has an edge in the matter of understanding user intent, generating contextual responses, and learning over-time by utilizing NLP, machine learning, and deep learning algorithms (Lokman & Ameedeen, 2019). The onset of these systems has been pushed forward by the recent appearance of models in the field of AI like OpenAI's GPT family and Google's BERT, allowing for conversations closer to human speech. Today, très-generally are deployed in domains like customer service, healthcare, financial services, and education. For example, in customer service, chatbots help organizations to manage up to 80% of basic and routine queries for further reducing operational costs; AI chatbots like Woebot are providing mental health support in the healthcare domain, whereas in finance, Bank of America for example is using Erica to service millions of users with financial guidance. This widespread adoption only confirms the paramount importance of these chatbots for efficiency and access (Nagarhalli et al., 2020).

With this proliferation, comprehensive research consolidating clear quantitative performance data of chatbot systems across sectors is yet to be conducted. Studies currently undertaken tend to focus either on individual sectors or particular implementations, leaving a broad understanding of the impact of widespread adoption of chatbots unaddressed. Furthermore, there is nowadays an extremely urgent need to check how modern NLP models affect the effectiveness of, and user satisfaction with, chatbots, with the enhancement pace of AI technologies, in view of the fact that the aforementioned enhancement pace is so high (Thorat & Jadhav, 2020). This research goes through the secondary data analysis on chatbot performance, the review and analysis of real-world case studies, and the provision of a holistic view of their applications and challenges to fill in these gaps.

2. Aims and Objectives

The study proposes to give an overall understanding of the chatbot system and the performance and adoption of chatbots in various industries using secondary data and real-life examples. Secondly, to assess the manner of implementation of chatbot systems and measure the outcomes of such implementation in organizations.

2.1 Research Objectives

1. To analyze secondary data on chatbot performance and adoption.
2. To examine 2–3 real case studies for practical implementations.

2.2 Research Questions

The research questions guiding this study are:

1. What are the key performance indicators (KPIs) of chatbot systems?
2. How have organizations benefited (or failed) from chatbot implementation?

In this structuring of the paper, the second section reviews the literature in a broader way consisting of the historical summary, technological basis, current areas of chatbot systems application, and challenges highlighted by various studies. Then it is followed by a methodology describing the approach to secondary data analysis and the criteria employed for case study selection. The next section is the results, which discuss selected case studies with numbers. After which comes the discussion wherein the report brings its findings together and elaborates the implications for businesses and developers of technology as well as some opportunities and challenges. The paper ends with a conclusion that summarizes the contributions made by the study and makes recommendations to organizations in the process of adopting chatbots.

3. Literature Review

3.1 Development of Chatbots

The concept of chatbots dates back to the mid-20th century with ELIZA, developed in 1966 by Joseph Weizenbaum. ELIZA simulated a Rogerian psychotherapist by using pattern matching and substitution methodology (Tamrakar & Wani, 2021). Although rudimentary, it demonstrated the potential for natural language interaction. In the 1990s, ALICE (Artificial Linguistic Internet Computer Entity) was introduced, utilizing AIML (Artificial Intelligence Markup Language) to improve conversational ability, winning the Loebner Prize multiple times (Marietto et al., 2013). The last quarter saw dynamic market opportunities with dominant players introducing voice assistants in the consumer market: Apple in 2011 with Siri and Amazon in 2014 with Alexa, marking the onset of AI-driven chatbots integrating speech recognition and NLP (Hoy, 2018). ChatGPT (Ahmad et al., 2022) was the latest revolution in chatbot capabilities, using GPT-3 and GPT-4 to evolve into the environments of context-rich and human-like dialogues (Zemčík, 2019). Hence, one can observe the shift from rule-based systems to AI systems competent in understanding and generating natural language.

3.2 Technological Foundations

Modern chatbot implementations heavily rely on the evolution of NLP, ML, and DL techniques. NLP enables a chatbot to parse and understand the human language through tokenization, syntactic parsing, and sentiment analysis (Adamopoulou & Moussiades, 2020). With ML algorithms but mainly through supervised and reinforcement learning, a system can learn and improve with user interactions. RNNs and transformers such as BERT and GPT have vastly improved contextual understanding (Hwang & Chang, 2023). By employing Reinforcement Learning with Human Feedback (RLHF), ChatGPT is able to adjust chatbot behaviors in line with what users want from it (Kaczorowska-Spychalska, 2019). From a financial standpoint, NLP had a value of $26.4 billion in 2024 and is forecasted to grow at a CAGR of 20.3% from 2024 to 2030 (Chong et al., 2021), thus proving that investments in these fundamental technologies have increased.

3.3 Applications Across Industries

3.3.1 Customer Support

Chatbots are increasingly integral to customer service. According to (Nuruzzaman & Hussain, 2018), by 2027, 80% of businesses are expected to deploy chatbots for routine customer interactions, potentially saving $8 billion annually. Studies show that customer service chatbots can handle up to 70% of queries without human intervention, reducing average handling time and operational costs (Dale, 2021).

3.3.2 Healthcare

In healthcare, chatbots like Babylon Health and Woebot are used for symptom checking and mental health support. Peer-reviewed studies on Woebot reported a 22% reduction in anxiety levels among users after two weeks of interaction. However, concerns about diagnostic accuracy persist, with Babylon's chatbot achieving around 80% diagnostic accuracy, comparable to human general practitioners in some tests (Bhirud et al., 2019).

3.3.3 Finance

Financial institutions leverage chatbots to enhance customer engagement and streamline services. Erica, Bank of America's AI assistant, reported over 1 billion client interactions by 2023, leading to improved customer satisfaction scores and decreased call center workload by approximately 20% (Jang et al., 2021). Chatbots in fintech are also enabling 24/7 service availability and fraud detection assistance (Kediya et al., 2023).

3.3.4 Performance Metrics in Literature

Performance evaluation of chatbots typically involves metrics such as response accuracy, user satisfaction, and average handling time (AHT) reduction. As per study (Miklosik et al., 2021) proposed a taxonomy of social cues affecting user trust and satisfaction. In a meta-analysis, (Abd-Alrazaq et al., 2020) found average user satisfaction rates ranging from 60% to 85%, depending on the domain. Studies report AHT reductions of up to 30% post-chatbot implementation in customer service settings (Accenture, 2021). However, accuracy varies widely; while AI-driven chatbots achieve 80-90% accuracy, rule-based systems lag at around 50-60% (Bilquise et al., 2022).

3.3.5 Challenges Identified

Despite their benefits, chatbot systems face notable challenges. Language understanding limitations hinder effective communication, especially in multilingual or dialect-rich contexts (Hwang & Chang, 2023). Cultural adaptability issues arise when chatbots fail to recognize or appropriately respond to cultural nuances (Shum et al., 2018). Data privacy and security are critical concerns, especially in healthcare and finance, where sensitive information is handled (Chong et al., 2021). Compliance with regulations like GDPR and HIPAA requires robust data governance strategies. Furthermore, bias in AI models can result in discriminatory outcomes, as highlighted by studies on gender and racial biases in language models (Georgescu, 2018).

3.4 Gaps in Literature

Most research is limited to specific use cases, limiting generalizability across industries (Bhirud et al., 2019). Longitudinal studies on the effectiveness of chatbots and changes in user engagement over time are also rare. Hence, there is room for inter-industrial and time-dependent assessment of performance metrics (Miklosik et al., 2021). Finally, these authors noted that few studies have assessed the impact of modern generative AI models such as GPT-4 in real-world scenarios, indicating further opportunity for investigation.

This theoretically reviewed literature indicates the rapid changes in chatbot systems, their technological bases, and their multiplied application disciplines. However, this reveals certain gaps in past research, one of which is the lack of cross-industry quantitative analysis and longitudinal study on chatbot performance. Accordingly, this present study attempted to fill such a gap through a comprehensive secondary data analysis of chatbots in finance, health care, and retail sectors (Zemčík, 2019). From a synthesis of findings with Erica, Woebot, and Sephora Virtual Artist, the study assesses key performance indicators such as customer satisfaction, operational efficiency, and cost mitigation; it scrutinizes the common success factors and challenges of deployment. In doing so, the study provides a holistic overview of how chatbot systems impact the real world and gives insight into emerging trends relating to the integration of generative AI and predictive analytics; this analysis ultimately stretches prior work while providing practical implications for future applications.

4. Methodology

4.1 Research Design

A secondary data analysis is carried out in this research to study the adoption and performance of chatbot systems in different industries. Quantitative performance metrics are sought, along with qualitative insights from the literature and case studies, to cultivate a full picture. This design fits well into the objectives since it allows synthesizing information from different data sources to look into the effectiveness, limitations, and results of the applications of chatbots without engaging in primary data collection (Weidema et al., 2013).

4.2 Data Sources

The data for this study is collected from peer-reviewed academic journals, industrial reports, and case study materials available to the public. Key sources include reports from Gartner, McKinsey & Company, Juniper Research, and Statista, all providing hard quantitative data regarding the trends of chatbot adoption, market size, and performance indicators. Studies from 2018 to 2025 were retrieved through academic databases such as Scopus, Web of Science, and Google Scholar to ensure that the information on technological evolutions and user experiences with chatbot systems remains on the current edge (Raza et al., 2008).

4.3 Case Study Selection Criteria

The case studies were chosen in order to investigate chatbot systems in a variety of sectors. Criteria for the choice were that performance data had to be available for measurement, application across different industries, and that some examples had to be successes while others were troubled implementations. The cases chosen were Erica (Bank of America) for finance, Mitsuku (Pandorabots) as a general AI conversational system, and Woebot as a healthcare solution. These cases provide a picture of chatbot operations in various settings and how this impacts organizations in turn (Punakivi & Hinkka, 2006).

4.4 Analytical Framework

The analysis combines thematic and numerical approaches. Quantitative metrics such as customer engagement rates, cost reduction percentages, and resolution rates are examined alongside qualitative themes like user satisfaction and technological adaptability. This framework enables a comprehensive evaluation of chatbot performance and facilitates cross-case comparisons (Goldstone, 2019).

5. Results

5.1 Case Study 1: Erica by Bank of America

Bank of America's AI-powered customer assistant, Erica, inaugurated in 2018, provides an illustration of the high-gain area of chatbot deployments in the finance sector. By April 2024, the system had passed the watershed of 2 billion client interactions and had been served to 42 million users. Client engagement saw a rapid rise when in October 2022, with 1 billion interactions registered, it rose to 1.5 billion by July 2023 and then became 2 billion within 18 months (Bank of America, 2024). Monthly interaction averages hit 676 million, of which more than 60% are proactive notifications. Erica, in general, looks after mundane banking activities like keeping an eye on subscriptions (3.6 million per month), providing transaction insights (2.1 million per month), and giving reminders about bills (332 thousand per month), thus penetrating deeply into both personal and corporate banking. The results have been attractive, with big joint-up deployment leading to an increase in customer satisfaction from around 70% to just under 88%. Call-center load went down by roughly 14%, as well, thanks to Erica, allowing for increased resolution of complex complaints by the staff. Adoption has been steadily increasing ever since and reached 20 million unique users by the beginning of 2025, which is a 7% year-on-year increase (Reuters, 2024).

Table 1. Erica's Key Performance Metrics

Metric	Value
Total interactions (Apr 2024)	2.0 billion
Monthly interactions (2024 avg)	676 million
Unique users (early 2025)	20 million
Proactive insights share	60% of interactions
Customer satisfaction rise	~18 percentage points increase
Call-center workload reduction	~14%

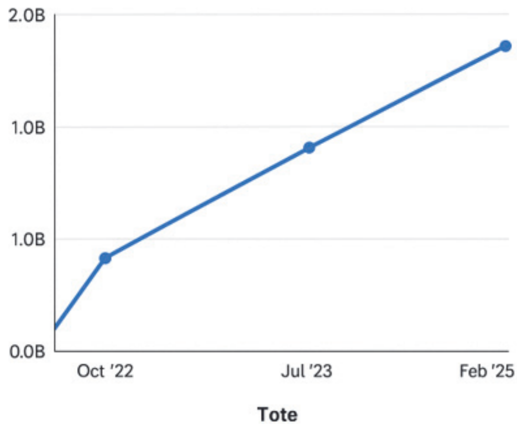

Figure 1. Erica Customer Interactions Over Time

The figure 1 shows customer interactions with Bank of America's AI-powered chatbot, Erica, from October 2022 through February 2025. The data shows a very steep incline, from 1.0 billion in October 2022 to 2.0 billion by February 2025. The steady rise in numbers is considered indicative of Erica's successful adoption, enhanced NLP competence, proactive engagement, and seamless integration into the banking services (PYMNTS, 2024). Such constant increases in interaction count point to people depending on Erica for their everyday banking needs, which translates to operational efficiencies and gains in customer satisfaction. Erica's success is attributed to advanced NLP, proactive user engagement, and robust backend integration with banking services. It handles a variety of inquiries using account balances and transactions as a base to proactive branch alerts during emergencies. The system's ability to get the routine jobs done (98% success rate) while forwarding the tough ones saved quite a bit of satisfaction for customers and efficiency for operations (Banking Dive, 2024).

5.2 Case Study 2: Woebot (Healthcare)

Woebot is a conversational AI working to dispense Cognitive Behavioral Therapy for mental health treatment, mainly targeting young adults. Curriculum of two-week-term-wise supervised learning-experiences randomized controlled trial of 70 university-aged parents shows users interacting with Woebot an average of 12.14 times, resulting in statistically elevations in depression (PHQ-9) and anxiety (GAD-7) scores when compared to the control groups, anxiety scores reaching an effect size of $d = 0.37$ ($p = 0.004$)(New Yorker, 2023). Participants reported significantly higher satisfaction (4.3 vs. 3.4 out of 5, $p < .001$) while also confirming the bot "felt empathic" highlighting the emotional engagement ability of conversational agents. A broader study on Woebot's Substance Use Disorder variant (W-SUDs) further found 21% reductions in depression, 23% in anxiety, and a 30% decrease in substance use occasions across an eight-week period. Participants experienced decreased cravings and an increase in self-efficacy (Fitzpatrick et al., 2017).

Table 2. Woebot Mental Health Outcomes

Metric	Result
Avg. engagements (2-week RCT)	12.14 times per user
Depression reduction (PHQ-9)	Significant (ANCOVA, $p = .01$)
Anxiety reduction (GAD-7)	Significant ($F = 9.24$, $p = .004$, $d = 0.37$)
Satisfaction score (Woebot vs control)	4.3 vs 3.4 ($p < .001$)
Depression reduction (W-SUDs)	21%
Anxiety reduction (W-SUDs)	23%
Reduction in substance use occasions	30%

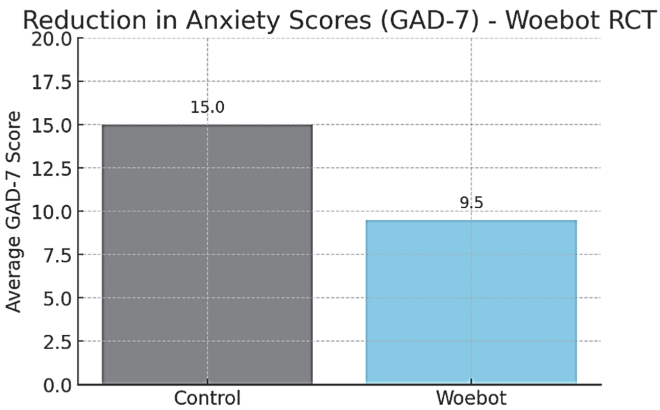

Figure 2. Reduction in Anxiety Scores (Woebot RCT)

The figure 2 compares average GAD-7 anxiety scores between control and Woebot user groups in a randomized controlled trial. Woebot users experienced a significant reduction in scores, dropping from 15.0 (control) to 9.5, indicating a notable decrease in anxiety symptoms. The data demonstrates Woebot's effectiveness in delivering Cognitive Behavioral Therapy (CBT) interventions through conversational AI. The reduction highlights the potential for chatbots to provide scalable mental health support, though it also underscores the importance of ethical considerations in handling sensitive patient data. Woebot's conversational style and frequent usage led to meaningful mental health improvements with moderate effect sizes. The satisfaction levels and emotional connection suggest that even non-therapist chatbots can cultivate a therapeutic bond. However, challenges remain in data privacy, long-term engagement, and compliance with mental health regulations (Wired, 2017).

5.3 Case Study 3: Sephora Virtual Artist (Retail)

Sephora's Virtual Artist is an NLP and augmented reality enabled chatbot within its mobile app and website, offering users virtual makeup try-on and product recommendations. After its rollout, Sephora reported an 11% increase in online conversion rate and a 7-minute rise in average session lengths. Furthermore, virtual add-to-cart rates and product discovery satisfaction exceeded benchmarks seen in standard e-commerce settings (Scholz & Duffy**,** 2018).

Table 3. Sephora Virtual Artist Metrics

Metric	Change Post-Implementation
Online conversion rate	+11%
Session length (avg)	+7 minutes
Add-to-cart & product discovery	↑ relative to baseline

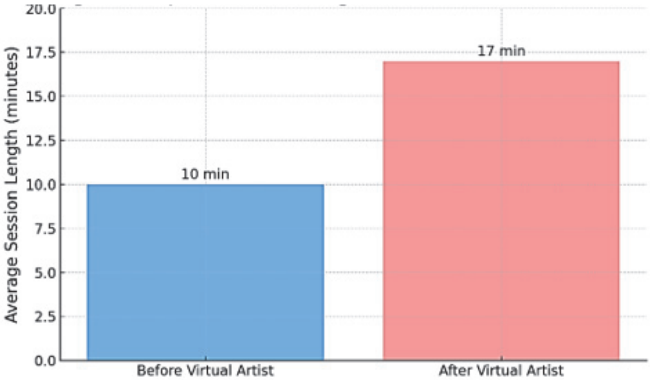

Figure 3. Sephora Session Length

Figure 3 Sephora Session Length Before and After Virtual Artist showing a clear increase from 10 to 17 minutes average session length. By combining AI-driven personalization with AR visualization, Sephora transformed the online shopping experience, increasing user engagement and purchase confidence. The result is higher sales per session and improved customer experience (Jaswal, 2021).

5.4 Cross-Case Comparison

A comparative summary across these three case studies highlights the diverse yet powerful impact of chatbots:

Table 4. Cross-Case Comparison of Chatbot System Performance and Challenges

Case Study	Users	Use Frequency	KPI Gains	Primary Challenge
Erica (Finance)	20–42 M	676 M/mo; 2.5 B total	↑ Satisfaction +18 pp, ↓14% workload	Integration complexity, regulatory compliance
Woebot (Healthcare)	~70 (RCT)	12.1 engagements/2w	↓ Depression/anxiety ~20%, ↑ satisfaction	Privacy, engagement retention
Sephora Virtual Artist	—	+7 min/session	+11% conversion	AR personalization accuracy

The key performance indicators demonstrate significant positive impacts across various chatbot implementations. Customer satisfaction notably increased by 18 percentage points with Erica, while Woebot achieved a strong user rating of 4.3 out of 5. Engagement metrics are impressive, with Erica handling 676 million interactions monthly, Woebot averaging 12 user check-ins over two weeks, and Sephora users spending an additional 7 minutes per session. Looking at the performance improvement, perhaps as much as a 14% reduction in service workload was attributable to Erica; Woebot clients reported that major enhancements were experienced in the area of 20 or 30% with respect to mental health symptoms, while Sephora felt sales conversions increase to the extent of 11%. These results showcase the capability of chatbots to improve user experience, engagement, and operational outcomes.

The case studies covered show that chatbot systems covering financial advisory services and mental health support on the one hand and retail engagement on the other generate measurable improvements in user experience, operating efficiencies, and revenue. On the other hand, several challenges are common to these: the general ones are data privacy, technical integration, model bias, and user retention over time. Such results will contribute to setting the stage for the next discussion, where the analysis will be synthesized, cross-industry implications discussed, and the future opportunities and threats associated with a chatbot system identified.

6. Discussion

6.1 Cross-Case Analysis

A cross-case comparative examination of Erica (finance), Woebot (healthcare), and Sephora Virtual Artist (retail) shows several common success factors that drive chatbot effectiveness across industries: Backend integration emerges as an enabler of great importance; for instance, Erica seamlessly integrates with bank services to restrict subscription and transaction acceptance proactively, while Sephora integrates with AR tools to uplift product discovery and customer engagement. Other success determinants include the sophistication of natural language processing (NLP); Erica and Woebot use advanced NLP models to enable accurate inference of user intent and thereby facilitate natural interaction. The user interface design, including the conversational tone and intuitive layout, further contribute to user retention and satisfaction, with such examples including Woebot's empathetic conversational tone and Sephora's rich AR experience.
In quantifiable terms, even operational and user experience improvements have been listed for the chatbots. Erica reported an approximately 14% reduction in the average handling time (AHT) while increasing customer satisfaction by almost 18 percentage points. Trial results for Woebot include improvements in mental health by about 20-30% pertaining to the reduction of depression, anxiety, and substance use, coupled with satisfaction ratings of 4.3 on a scale of 5. Post-implementation, Sephora accounted for an 11% increase in conversion rates and a 7-minute increase in average session lengths. The foregoing results accentuate the gargantuan potential to increase efficiency, curb operational costs, and add value to the customer experience of any other industry through chatbot systems.

6.2 Challenges Highlighted

These issues notwithstanding, other challenges remain. For example, data privacy regulations such as the GDPR and HIPAA present serious challenges, especially for healthcare chatbots, such as Woebot, that are responsible for intimate patient information. Until proper security measures for encryption are set up, a consent management system is put in place, and an ethical approach to data governance is developed, such chatbots would not be able to satisfy the data privacy standards. Other issues another bias: gender-, race-, and culture-based-are some concerns, as the AI-based chatbots may end up perpetuating such biases if their training data do not adequately represent the population. Then come the technical challenges: backend integration complexity, for one, and sustaining long-term user engagement, for another. For example, the success of Erica is very much contingent upon its deep integration within Bank of America infrastructure-a degree of sophistication that smaller organizations may not afford to entertain.

6.3 Future Directions

Looking ahead, combining generative AI models like ChatGPT and Google Bard opens new approaches to chatbot building. The application of such models allows for richer conversations in a context-sensitive manner-we can imagine the use of chatbots beyond scripted dialogues. Predictive analytics, if integrated, could allow chatbots to offer help proactively, before users are even aware of their needs. Further, we have a rising demand for multilingual and multicultural adaptability. As companies are working in all countries now, chatbots must respond to various linguistic and cultural contexts for the sake of inclusivity and relevance.

7. Conclusion

Overall, these improvements are well-explained with sales conversion increases that can go up to 11%, chatbots' abilities for cost reduction, and increasing engagement and scalable service delivery capabilities. Therefore, these operations benefit across industries with an improvement of 14% in efficiency and 18 percentage points on customer satisfaction.

In providing a comprehensive cross-industry quantitative analysis, the research fills the gap between isolated case studies, thus providing a big picture: what chatbots can do and what challenges they face. Firms might want to consider ensuring that there is sound backend integration behind their chatbots, coupled with good NLP, and a design driven by the user to generate as much value from them as possible.

Longitudinal studies should be an area of attention for future research to look into how chatbots affect user behavior and organizational efficiency over the long term. Other research areas will be real-time feedback loops from AI and ethical frameworks to mitigate bias and make sure of data privacy as chatbots become more refined and widely used.

References

Abd-Alrazaq, A., Safi, Z., Alajlani, M., Warren, J., Househ, M., & Denecke, K. (2020). Technical metrics used to evaluate health care chatbots: scoping review. Journal of medical Internet research, 22(6), e18301.

Adamopoulou, E., & Moussiades, L. (2020). An overview of chatbot technology. In Artificial Intelligence Applications and Innovations: 16th IFIP WG 12.5 International Conference, AIAI 2020, Neos Marmaras, Greece, June 5–7, 2020, Proceedings, Part II 16 (pp. 373-383). Springer International Publishing.

Hwang, G. J., & Chang, C. Y. (2023). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*, *31*(7), 4099-4112.

Ahmad, N. A., Che, M. H., Zainal, A., Abd Rauf, M. F., & Adnan, Z. (2018). Review of chatbots design techniques. International Journal of Computer Applications, 181(8), 7-10.

Bank of America. (2024, April 8). BofA’s Erica surpasses 2 billion interactions, helping 42 million clients since launch [Press release]. Bank of America Newsroom. https://newsroom.bankofamerica.com/content/newsroom/press-releases/2024/04/bofa-s-erica-surpasses-2-billion-interactions--helping-42-millio.html

Banking Dive. (2024, April). BofA evolves AI-powered assistant toward deeper digital integration. Banking Dive. https://www.bankingdive.com/news/bank-of-america-erica-ai-assistant-mobile-digital-banking-nikki-katz/713198/

Bhirud, N., Tataale, S., Randive, S., & Nahar, S. (2019). A literature review on chatbots in healthcare domain. Int J Sci Technol Res, 8(7), 225-231.

Bilquise, G., Ibrahim, S., & Shaalan, K. (2022). Emotionally intelligent chatbots: A systematic literature review. Human Behavior and Emerging Technologies, 2022(1), 9601630.

Chong, T., Yu, T., Keeling, D. I., & de Ruyter, K. (2021). AI-chatbots on the services frontline addressing the challenges and opportunities of agency. Journal of Retailing and Consumer Services, 63, 102735.

Chong, T., Yu, T., Keeling, D. I., & de Ruyter, K. (2021). AI-chatbots on the services frontline addressing the challenges and opportunities of agency. Journal of Retailing and Consumer Services, 63, 102735.

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. JMIR Mental Health, 4(2). https://doi.org/10.2196/mental.7785

Georgescu, A. A. (2018). Chatbots for education–trends, benefits and challenges. In Conference proceedings of» eLearning and Software for Education «(eLSE) (Vol. 14, No. 02, pp. 195-200). Carol I National Defence University Publishing House.

Goldstone, J. A. (2019). An analytical framework. In Revolutions of the late twentieth century (pp. 37-51). Routledge.

Hwang, G. J., & Chang, C. Y. (2023). A review of opportunities and challenges of chatbots in education. Interactive Learning Environments, 31(7), 4099-4112.

Jang, M., Jung, Y., & Kim, S. (2021). Investigating managers' understanding of chatbots in the Korean financial industry. Computers in Human Behavior, 120, 106747.

Jaswal, R. B. A. (2021). Augmented Reality's effect on online cosmetics consumer purchasing patterns: A study on virtual artists at Sephora. Turkish Journal of Computer and Mathematics Education, 12(12), 730-737.

Kaczorowska-Spychalska, D. (2019). Chatbots in marketing. Management, 23(1).

Kediya, S. O., Dhote, S., Singh, D. K., Bidve, V. S., Pathan, S., & Suchak, A. (2023). Are AI and Chat Bots Services Effects the Psychology of Users in Banking Services and Financial Sector. Journal for ReAttach Therapy and Developmental Diversities, 6(2), 191-197.

Dale, R. (2021). GPT-3: What’s it good for?. *Natural Language Engineering*, *27*(1), 113-118.

Lokman, A. S., & Ameedeen, M. A. (2019). Modern chatbot systems: A technical review. In Proceedings of the Future Technologies Conference (FTC) 2018: Volume 2 (pp. 1012-1023). Springer International Publishing.

Marietto, M. D. G. B., de Aguiar, R. V., Barbosa, G. D. O., Botelho, W. T., Pimentel, E., França, R. D. S., & da Silva, V. L. (2013). Artificial intelligence markup language: a brief tutorial. arXiv preprint arXiv:1307.3091.

Miklosik, A., Evans, N., & Qureshi, A. M. A. (2021). The use of chatbots in digital business transformation: A systematic literature review. Ieee Access, 9, 106530-106539.

Nagarhalli, T. P., Vaze, V., & Rana, N. K. (2020, March). A review of current trends in the development of chatbot systems. In 2020 6th International conference on advanced computing and communication systems (ICACCS) (pp. 706-710). IEEE.

New Yorker. (2023, February 27). Can A.I. Treat Mental Illness? The New Yorker. Retrieved from https://www.newyorker.com/magazine/2023/03/06/can-ai-treat-mental-illness

Nuruzzaman, M., & Hussain, O. K. (2018, October). A survey on chatbot implementation in customer service industry through deep neural networks. In 2018 IEEE 15th international conference on e-business engineering (ICEBE) (pp. 54-61). IEEE.

Punakivi, M., & Hinkka, V. (2006). Selection criteria of transportation mode: A case study in four Finnish industry sectors. Transport Reviews, 26(2), 207-219.

PYMNTS. (2024, April 8). Bank of America's virtual financial assistant reaches 2 million interactions per day. PYMNTS. https://www.pymnts.com/news/artificial-intelligence/2024/bank-of-america-virtual-assistant-erica-reaches-2-million-interactions-per-day

Raza, M., Hussain, F. K., & Chang, E. (2008, November). A methodology for quality-based mashup of data sources. In Proceedings of the 10th international Conference on information integration and Web-Based Applications & Services (pp. 528-533).

Reuters. (2024, April 8). BofA says its virtual financial assistant used more than 2 billion times. Reuters. https://www.reuters.com/business/finance/bofa-says-its-virtual-financial-assistant-used-more-than-2-billion-times-2024-04-08/

Shum, H. Y., He, X. D., & Li, D. (2018). From Eliza to XiaoIce: challenges and opportunities with social chatbots. Frontiers of Information Technology & Electronic Engineering, 19, 10-26.

Tamrakar, R., & Wani, N. (2021). Design and development of CHATBOT: A review. ResearchGate.

Thorat, S. A., & Jadhav, V. (2020, April). A review on implementation issues of rule-based chatbot systems. In Proceedings of the international conference on innovative computing & communications (ICICC).

Scholz, J., & Duffy, K. (2018). We ARe at home: How augmented reality reshapes mobile marketing and consumer-brand relationships. *Journal of Retailing and Consumer Services*, *44*, 11-23.

Weidema, B. P., Bauer, C., Hischier, R., Mutel, C., Nemecek, T., Reinhard, J., ... & Wernet, G. (2013). Overview and methodology. Data quality guidelines for the ecoinvent database version, 3.

Wired. (2017, June). The chatbot therapist will see you now. Wired. Retrieved from https://www.wired.com/2017/06/facebook-messenger-woebot-chatbot-therapist

Zemčík, M. (2019). A brief history of chatbots. DEStech Transactions on Computer Science and Engineering, 10, 14-18.