Issues in Information Systems
Volume 26, Issue 1, pp. 446-457, 2025

DOI: https://doi.org/10.48009/1_iis_133

From human teller to virtual assistant: A study on chatbot adoption in leading banks

Wen-Hsiu Chou, *Florida International University, wchou@fiu.edu*
Chen-Huei Chou, *College of Charleston, chouc@cofc.edu*

Abstract

Artificial Intelligence has developed steadily over recent decades. Because of the swift adoption across various applications, particularly generative AI, AI techniques have been adopted in many industries. One prominent application is the AI-powered chatbot which is commonly used to improve operational efficiency, enhance customer service, personalize and tailor responses, improve cost effectiveness, and so on. This study aims to evaluate AI chatbots implemented by top banks in the United States. Systematic and comprehensive analyses are conducted on these AI chatbots using the same pre-designed tasks. The chatbots are evaluated based on their precision, accuracy, and failure. Advantages and disadvantages of these AI chatbots are identified. Suggestions for the future improvements of these chatbots are provided. Overall findings suggest that the integration of AI into bank chatbots should not be seen as a binary shift from human to machine, but rather as a complementary tool that enhances, yet does not replace human capital.

Keywords: virtual assistant, chatbot, and banking industry

Introduction

The integration of artificial intelligence (AI) into a wide range of industries has gained momentum in recent years, with chatbots emerging as a key interface for consumer interaction and service delivery. The release of ChatGPT-4o by OpenAI on May 13, 2024, represents a significant milestone in the ongoing evolution. With real-time reasoning capabilities across text, audio, and vision, ChatGPT-4o pushes the frontier of what AI-driven conversational agents can accomplish. The benefits of using AI chatbots include increased efficiency, enhanced customer service, personalized and tailored responses, improved cost effectiveness, and so on. To the best of our knowledge, little research has been conducted on retail banking services that cater to the everyday financial needs of individuals.

This study aims to be one of the first to evaluate AI chatbots deployed in the top banks (Bond, 2024) in the United States, based on their assets. The banking industry, particularly in the United States, is a sector characterized by strict regulatory oversight to ensure financial stability and protect consumers. In response to evolving consumer expectations and competitive pressures, U.S. banks have made substantial investments in technological development. This unique framework motivates the following research questions: Are leading U.S. banks leveraging cutting-edge AI technologies to enhance service delivery? Does current AI adoption contribute to the substitution of human labor with virtual assistants?

Table 1 summarizes the characteristics of the top banks in terms of the rank, type of bank (traditional vs. online), founded year, assets, number of branches, focus, unique features, Chatbot name, and Chatbot description. The description for the JP Morgan Chase chatbot was taken from https://www.jpmorgan.com/payments/solutions/access/jpm-virtual-assistant, Erica from Bank of America was taken from https://promotions.bankofamerica.com/digitalbanking/mobilebanking/erica and https://sites.wf.com/fargo/ was used for Wells Fargo.

Table 1. Characteristics of top banks in the United States.

Name of Bank	JPMorgan Chase & Co	Bank of America	Citibank	Wells Fargo	Ally Bank
Rank	1 in Traditional	2 in Traditional	3 in Traditional	4 in Traditional	3 in Online
Type	Traditional	Traditional	Traditional	Traditional	Online
Year	1799	1784	1812	1852	2009 (before: GMAC bank, established in 2000)
Assets	$3.58 trillion	$2.56 trillion	$1.73 trillion	$1.69 trillion	$182.1 billion
Branches	More than 4,900 across 48 states and Washington D.C.	More than 3,700 across 40 states and Washington D.C.	More than 600 across 13 states and Washington D.C.	More than 4,200 across 36 states and Washington D.C.	None
Focus	Retail and commercial banking	Retail and wealth management	Global consumer banking		Auto loans and online savings accounts
Unique Features	Biggest bank by assets	Leader in digital banking technology	Leading global bank, offering financial services in 180 different countries		Very competitive interest rates
Chatbot Name	Chase digital assistant	Erica (virtual financial assistant)	Citi® Bot	Fargo®	Ally Assist
Chatbot description	J.P. Morgan Virtual Assistant can assist customers with tasks such as tracking wire transfers and retrieving account balance and transaction reports.	Erica® is a virtual financial assistant in the Bank of America Mobile Banking app. It provides personalized insights, helps with transactions, and connects users to specialists when needed.	Citi® Bot mainly focuses on account questions with the topics such as transfers between bank accounts, linking accounts, credit card payments, etc.	Fargo® is embedded in the Wells Fargo Mobile® app. It assists users with everyday banking needs, including checking balances, reviewing transactions, and providing spending insights.	Ally Assist helps users manage their finances by providing real-time account balances and performing tasks such as making payments, transferring funds, and setting up alerts.

Literature Review

Wu (2024) provides the development of chatbot technology and examines how chatbot, specifically IBM's Watson Assistant, affects bank performance by improving cost efficiency, customer satisfaction, and staff productivity. The paper also highlights a concern regarding perceived risk in the use of chatbots which includes both risk perception and uncertainty from the user's perspective. Chatbot systems can contribute to perceived risk when customers feel that their problems are not adequately understood or addressed, potentially leading to diminished trust in the bank's services. Wube et al. (2022) conducted a systematic review of text-based chatbots implementation in the financial sector, highlighting their role in improving user engagement, trust, and addressing security concerns. Yatawara et al. (2025) further provide a comprehensive literature review across sectors, including banking, and identify critical factors that affect consumer adoption of AI-driven chatbots: user experience, technological quality, and behavioral factors, which form the foundation for our bank chatbots evaluation.

Empirical studies confirm the relevance of service quality in financial chatbots. For example, Graham et al. (2025) explore the transformative role of chatbots in banking and financial services, arguing that these AI tools are not merely cost-saving innovations but potential catalysts for a broader organizational shift toward AI-enabled service models. Qureshi et al. (2024) examine the influence of chatbot service quality on customer satisfaction in banking and identify responsiveness and reliability as critical determinants. Likewise, the Consumer Financial Protection Bureau (CFPB, 2023) reported that the majority of banks now deploy chatbot technologies, yet challenges in conversational accuracy and oversight limit effectiveness. In a financial context, accuracy alone is insufficient as customers also value contextual appropriateness and correct resolution of queries. Thus, our dual-focus design (precision + correct assistance rate) aligns with best practices from both the operational research literature and behavioral adoption research.

To maintain sustainable growth and competitiveness, making significant investments in AI technology and adapting to changing market dynamics has become an inevitable trend for financial institutions (Zhao, Tsai, and Wang, 2019). Based on 522 valid survey responses, Barjaktarovic Rakocevic et al. (2025) find that customers identified digital banking services as the most important factor when choosing a bank, ranking it above branch proximity or lending conditions. In addition, Bhatnagr, Rajesh, and Misra (2024) indicate that a good interaction leads to higher satisfaction, better user experience, and a stronger willingness to keep using AI-powered digital banking, which means AI should be designed to be easy to use, fast, and dependable. These features are essential to ensuring a bank's long-term success and sustainability. Our study aims to fill the gap in understanding the current adoption of chatbots among leading banks in the United States and contributes to both academia and practice by providing valuable insights on digital transformation in the banking industry, which is particularly important and has drawn significant attention following the launch of ChatGPT.

Using the Technology Acceptance Model (TAM), Alt et al. (2024) use the Structural Equation Modeling method to analyze the factors for the intention to use a banking chatbot. They found that the perceived usefulness of the chatbot and perceived compatibility directly impacted the intention to use it. Additionally, they found that awareness of the chatbot service indirectly influenced the intention to use a banking chatbot. Since banking chatbots are normally proprietary assets, the technologies are not disclosed. Graham et al. (2025) tailored interview questions by following the Unified Theory of Acceptance and Use of Technology (UTAUT) model to examine the use of chatbots in customer service within the banking and finance sector. Their results are summarized based on the aspects of UTAUT. For performance expectancy, chatbots are instant, while human advisors are slow. Security and risk management are the measures of effort expectancy. Concerns are that banks may not secure the data being transferred to a third-party platform. For social

influence, chatbots should have a personality to be likeable and enjoyable. They also found that chatbots' capabilities are limited and human assistance is still needed to complete customer queries.

Adamopoulou and Moussiades (2020) summarize the fundamental chatbot technologies as pattern matching, latent semantic analysis, chatscript, rivescript, natural language processing, and natural language understanding. Shahriar and Hayawi (2023) identify several domains that have adopted ChatGPT, including medicine and public health, education, reasoning, journalism and misinformation detection, software development, translation, and scientific research. Banking chatbots were not identified as one of the domains. In this study, we evaluate the chatbots deployed by the top banks in the United States.

In our study, we compare chatbots used by online-only banks with those used by banks that operate physical branches. It is worth noting that we do not include Morgan Stanley Private Bank and American Express in our results, despite their ranking as the largest and second-largest online-only U.S. banks by assets as of December 2024. Morgan Stanley Private Bank primarily focuses on wealth management and investment services rather than traditional retail banking, and therefore falls outside the scope of our research objectives.

As for American Express, we interacted with its chatbot; however, human intervention occurred immediately after we asked the first basic question related to bank accounts. This suggests that American Express may intentionally limit chatbot autonomy to avoid potential credibility risks if customer inquiries are misunderstood. Using panel data from 2010 to 2020 across 20 countries, Gyau et al. (2024) show that finance AI technology innovation has a positive impact on banks' return on assets (ROA), but the benefits of AI appear strongest in the first two years after adoption but diminish over time due to rising implementation costs or market saturation. The following section describes the methodology.

Methodology

We conducted comprehensive analyses in systematic comparisons among the AI chatbots using the same pre-designed tasks. Structured or pre-defined sequences of interactions used in the experiment help standardize the data collection process, allowing us to perform effective comparisons across different platforms. Chalyi (2024) employs confusion matrices and a pairwise comparison methodology to assess chatbots against eight specific criteria. In the medical domain, researchers have adopted confusion-matrix-driven evaluation to assess chatbots' factual accuracy and reliability. For example, Zúñiga Salazar et al. (2023) evaluate ChatGPT, Google Bard, and Bing AI on triaging medical questions (emergency vs. non-emergency) and find that Google Bard detected 87% of actual emergency cases and 36% of non-emergencies (true positive and true negative rates, respectively). They demonstrate the use of domain-specific ground truth data and confusion metrics to rigorously compare chatbot performance in a safety-critical application.

Similarly, Lee et al. (2024) compare several large language model chatbots on a medical board exam question bank, using the percentage of correct answers as the metric, and show that confusion-matrix based measures can quantify chatbot knowledge performance. These studies reinforce that precision/accuracy metrics are vital for evaluating chatbot in Q&A-style assessments. In our study, to evaluate chatbot responses from the selected banks, we followed Chalyi's method and used metrics such as accuracy rate and precision to evaluate the chatbots. We utilized a variety of pre-defined tasks in 10 different categories, as listed below, with a planned 30 questions to evaluate the chatbots. Each category consisted of two to four related questions. The data collected are the responses and directions provided by the chatbots. The next section shows the pre-defined tasks.

Pre-defined Tasks

1. Checking Accounts
    a. What are the requirements to open a checking account?
    b. What are the monthly maintenance fees of the checking account? Is there any way I can waive them?
    c. How can I order checks for my account?
    d. What is the daily withdrawal limit for a checking account?

2. Savings Accounts
    a. What is the current interest rate for savings accounts?
    b. Are there any fees associated with maintaining a savings account?
    c. Can I set up automatic transfers from my checking to my savings account?
    d. Is there a limit of withdrawals I can make from a savings account?

3. Certificates of Deposit (CDs)
    a. What are the terms and interest rates for CDs currently available?
    b. What is the minimum deposit required to open a CD?
    c. Can I roll over a CD automatically when it matures?
    d. Can I withdraw funds from a CD before maturity, and are there penalties?

4. Credit Cards
    a. What credit cards do you have available? What are their characteristics in terms of annual fees and rewards?
    b. What are the eligibility requirements for your credit cards?
    c. How long does it take to get approved for a credit card?
    d. Can I apply for a credit card online or through the mobile app?

5. Mortgages
    a. What are the current mortgage interest rates?
    b. What documents do I need to apply for a mortgage?
    c. Can I calculate my estimated monthly payments before applying?

6. Special Promotions
    a. Are there any promotions currently available for opening a new checking or savings account?
    b. Do you have promotional credit card offers?
    c. Are there any seasonal promotions I should be aware of?

7. Suggestions
    a. Help me choose a credit card that is right for me
    b. What's the best way to save money using your banking services?

8. Customer Support
    a. How do I report a lost or stolen card?
    b. Can you help me dispute a transaction on my account?

9. Mobile and Online Banking
    a. Can I open an account from here?
    b. Can I receive account alerts or notifications?

10. Credit Score and Financial Health
    a. Do you provide tools to help monitor or improve my credit score?
    b. What financial wellness programs do you offer?

Evaluation Metrics

In this study, we used the following metrics to evaluate the chatbots used by the banks. The metrics include precision, accuracy rate, failure, true positive, false positive, true negative, and false negative. Here are the definitions of these metrics:

- True Positive (TP): the number of correct perceived understandings by a chatbot of our pre-defined question when the interaction or guidance is correct.
- False Negative (FN): the number of correct perceived understandings by a chatbot of our pre-defined question when the interaction or guidance is not correct.
- True Negative (TN): although "the number of incorrect perceived understandings by a chatbot of our pre-defined question when the interaction or guidance is not correct" seems to be applicable, it does not apply in the context of the experiment. The reason is that if a chatbot can't comprehend a question, providing an off-target answer is not a correct prediction.
- False Positive (FP): the number of incorrect perceived understandings by a chatbot of our pre-defined question when the interaction or guidance is correct.
- Precision=$\frac{TP}{TP+FP}$: the precision rate shows the percentage of actual understandings out of all correct perceived understandings by a chatbot.
- Failure is the case that the Chatbot won't be able to comprehend the meaning of the question, so that no correct or related guidance is provided.
- Accuracy rate= $\frac{TP}{TP+FP+FN}$: the accuracy rate shows the percentage of correct perceived understandings by a chatbot out of all cases. The original accuracy rate is $\frac{TP+TN}{TP+TN+FP+FN}$. Since TN is not used, the accuracy rate in this study is $\frac{TP}{TP+FP+FN}$.

Inter-rater Reliability

When there are multiple raters evaluating the given tasks, Cohen's Kappa (Cohen, 1960) is normally used to validate the inter-rater reliability. In this study, two raters evaluate each chatbot's performance in its binary results, either pass or fail, on the perceived understanding and provided assistance. As a result, the Cohen's Kappa (k) is calculated as follows:

$$k = \frac{2 \times (TP \times TN - FN \times FP)}{(TP + FP) \times (FP + TN) + (TP + FN) \times (FN + TN)}$$

Results

All the selected banks underwent a list of 30 questions in 10 categories to evaluate their chatbots. Each category consists of two to four questions. Following the evaluation metrics mentioned above, two raters evaluated each chatbot's performance in terms of perceived understanding and provided assistance. Table 2 shows the first-round result of Cohen's Kappa. Based on Landis and Koch's suggestion (1977), the agreement is almost perfect when the Cohen's Kappa is above 0.8. When a value is between 0.6 and 0.8, the agreement is substantial. To make the evaluations of chatbots more manageable, in the second round the raters eventually reached a consensus on judging the outcomes generated by the chatbots. The remaining experiments are based on the agreed-upon performance outcomes for evaluation.

Table 2. Cohen's Kappa for two raters evaluating the results provided by chatbots in the United States.

	Chase	Bank of America	Citi	Wells Fargo	Ally
Perceived Understanding by chatbot	86.77%	85.65%	91.8%	86.4%	92.96%
Provided assistance by chatbot	88.64%	78.02%	84.13%	86.11%	85.65%

Table 3 reports the precision of the chatbots' performance. All chatbots reached 100% in precision in most categories and provided an overall 100% precision as well. This is due to the fact that there were no False Positive cases made in chatbots, meaning that there were no chatbots misunderstood a question while providing a correct interaction or guidance.

Table 3. Precision of top banks' chatbots in the United States.

Category	Chase	Bank of America	Citi	Wells Fargo	Ally
Checking Accounts	100%	100%	100%	100%	100%
Savings Accounts	100%	100%	100%	100%	100%
Certificates of Deposit (CDs)	N/A	100%	100%	100%	100%
Credit Cards	100%	100%	100%	100%	100%
Mortgages	100%	N/A	100%	100%	100%
Special Promotions	N/A	100%	100%	100%	100%
Suggestions	N/A	N/A	100%	100%	100%
Customer Support	100%	100%	100%	100%	100%
Mobile and Online Banking	100%	100%	100%	100%	100%
Credit Score and Financial Health	100%	100%	100%	100%	100%
Overall	100%	100%	100%	100%	100%

Table 4 presents the accuracy rates of the chatbots across the same 30 questions, categorized into 10 categories. Overall, Citi Bot and Fargo were the top chatbots that reached an accuracy rate of 95%. Chase and Ally also followed as the 3rd and 4th performers with an accuracy rate over 90%. Bank of America's Erika was the least accurate chatbot in comparison, generating only 74% correct predictions.

Table 4. Accuracy of top banks' chatbots in the United States.

Category	Chase	Bank of America	Citi	Wells Fargo	Ally
Checking Accounts	100%	67%	100%	100%	100%
Savings Accounts	100%	75%	100%	100%	100%
Certificates of Deposit (CDs)	N/A	100%	100%	100%	100%
Credit Cards	100%	50%	100%	75%	N/A
Mortgages	100%	0%	100%	100%	0%
Special Promotions	0%	100%	50%	100%	N/A
Suggestions	N/A	0%	100%	100%	N/A
Customer Support	100%	100%	100%	100%	100%
Mobile and Online Banking	100%	100%	100%	100%	100%
Credit Score and Financial Health	100%	100%	100%	100%	N/A
Overall	93%	74%	95%	95%	91%

Although we can see quite high precision and accuracy rates in Tables 3 and 4, these two measures do not capture the cases where a chatbot couldn't understand the given question, so that an unrelated interaction or guidance was provided. Table 5 summarizes the total number of failures made by the chatbots. Additionally, Table 6 is updated to reflect the correct assistant rates for the chatbots. In general, tasks related to "Certificates of Deposit" were the most challenging for chatbots, while "Customer Support" related tasks were handled flawlessly by the chatbots.

Table 5. Failure cases of top banks' chatbots in the United States.

Category	Chase	Bank of America	Citi	Wells Fargo	Ally
Checking Accounts	3	1	0	1	1
Savings Accounts	1	0	0	2	1
Certificates of Deposit (CDs)	4	3	3	3	3
Credit Cards	2	2	2	0	4
Mortgages	1	2	1	2	2
Special Promotions	2	1	1	1	3
Suggestions	2	1	0	1	2
Customer Support	0	0	0	0	0
Mobile and Online Banking	0	0	0	0	1
Credit Score and Financial Health	0	1	1	1	2
Overall	15	11	8	11	19

Table 6. Correct assistant rate of top banks' chatbots in the United States.

	Chase	Bank of America	Citi	Wells Fargo	Ally
Checking Accounts	25%	50%	100%	75%	75%
Savings Accounts	75%	75%	100%	50%	75%
Certificates of Deposit (CDs)	0%	25%	25%	25%	25%
Credit Cards	50%	25%	50%	75%	0%
Mortgages	67%	0%	67%	33%	0%
Special Promotions	0%	67%	33%	67%	0%
Suggestions	0%	0%	100%	50%	0%
Customer Support	100%	100%	100%	100%	100%
Mobile and Online Banking	100%	100%	100%	100%	50%
Credit Score and Financial Health	100%	50%	50%	50%	0%
Overall	47%	47%	70%	60%	33%

Citibank's Citi Bot is the overall best, as it achieved the highest accuracy rate of 95%, the lowest failure cases (8 out of 30 questions), and the highest correct assistance rate of 70%. Wells Fargo's Fargo chatbot is the second best with an accuracy rate of 95%. It failed in 11 out of 30 cases and correctly assisted 60% of cases. Bank of America's Erika is the worst chatbot provided by traditional banks in this study, as its accuracy rate was 74%, failing in 11 cases and correctly assisting 47% of cases. Although Ally's chatbot's accuracy was 91%, it failed in 19 cases and correctly assisted in only 33% of cases.

Discussions and Future Directions

This study is limited to a select group of top-ranked banks. The limitation was also due to the bank's access restriction to its existing customers. Without being their customers, their chatbot can't be used. Future researchers are encouraged to become the customers of all top-ranked banks to conduct a more comprehensive comparison of chatbots' performance. Surprisingly, Ally, as the only online bank in the comparison, doesn't offer a decent chatbot, unlike traditional banks that also provide online banking services.

Understanding customer queries by chatbots is only a fundamental capability, but there remains substantial room for improvement. While AI has improved access to services in offering 24/7 availability, its capacity for meaningful engagement, such as interpreting customer needs, conducting multi-layered bank product comparisons, or offering tailored financial advice is still limited.

In many cases, chatbots merely serve as front-line filters, redirecting users to predefined FAQs or bank websites. For example, for a customer query "What are the monthly maintenance fees of the checking account? Is there any way I can waive them?", a chatbot may redirect the user to a webpage or ask the user to select the account the user has in order to proceed. A chatbot may not be able to get the meaning of the question by providing a way to avoid fees for checking, saving, and credit card accounts. Another similar example, such as "What is the current interest rate for savings accounts?", a chatbot may direct users to a page with all types of accounts without directly answering the interest rate question. In addition, "CD" is the abbreviation for Certificate of Deposit. Interestingly, most chatbots cannot comprehend the meaning of "CD" in related questions, often providing unrelated responses and guidance. Also, chatbots often fail to engage in reasoning or interpret contextually complex queries. As such, the enhancement in service delivery is largely operational rather than strategic or cognitive. Bank chatbots, in their current form, are best viewed as a complement to rather than a substitute for human labor. At the current stage of development, human expertise remains indispensable, as AI technologies are still far from serving as a substitute for human judgment.

Conclusion

To ensure long-term growth and remain competitive in an increasingly digital economy, financial institutions like banks are compelled to make strategic and sustained investments in artificial intelligence (AI). This shift is no longer optional but a necessary response to rapidly evolving market dynamics and changing consumer expectations. This study contributes to the growing literature on AI in financial services by systematically evaluating the adoption and performance of AI-powered chatbots in leading U.S. banks. In a sector marked by regulatory complexity and consumer trust considerations, our analysis reveals that chatbot deployment among top-tier banks remains centered on transactional support, basic inquiries, and customer engagement.

Our findings highlight that current AI chatbot systems adopted by banks are far from replacing human customer service agents. Rather, they function as complementary tools that enhance efficiency, reduce routine workloads, and provide 24/7 accessibility. While generative AI models like ChatGPT-4o have expanded the technical frontier—enabling real-time reasoning and multimodal capabilities—their full integration into bank service platforms remains nascent. Most chatbots still rely on rule-based or narrowly trained models with limited contextual flexibility.

Our comparative evaluation across leading U.S. banks reveals both encouraging advancements and persistent limitations in chatbot performance. Citibank's Citi Bot emerged as the overall best performer, achieving the highest accuracy rate of 95%, the fewest failure cases (8 out of 30), and the highest correct assistance rate of 70%. Wells Fargo's Fargo followed closely with a similarly high accuracy rate of 95%, though it failed in 11 instances and provided correct assistance in 60% of cases. In contrast, Bank of America's Erica, despite the bank's strong digital reputation, was the weakest among traditional banks in this study. Erica recorded a notably lower accuracy rate of 74%, failed in 11 cases, and achieved a correct assistance rate of only 47%. Meanwhile, Ally Bank's chatbot, though demonstrating a respectable 91% accuracy rate, exhibited a high number of failures (19 out of 30) and the lowest correct assistance rate at just 33%.

These findings underscore that technological sophistication alone does not guarantee consistent chatbot performance. Common limitations—such as difficulty handling multi-intent queries, or offering overly generic responses—continue to impede user satisfaction and highlight areas for further improvement in natural language understanding and contextual reasoning. Thus, banks that invest early and integrate AI thoughtfully are more likely to secure competitive advantages and remain resilient amid ongoing digital disruption.

References

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.

Alt, M. A., Vizeli, I., & Săplăcan, Z. (2021). Banking with a chatbot – A study on technology acceptance. *Studia Universitatis Babes-Bolyai, Oeconomica*, *66*(1).

Barjaktarovic Rakocevic, Sladjana, Nela Rakic, and Rade Rakocevic. (2025). An Interplay Between Digital Banking Services, Perceived Risks, Customers' Expectations, and Customers' Satisfaction. *Risks* 13: 39.

Bhatnagr, Puneett, Anupama Rajesh, and Richa Misra. (2024). Continuous intention usage of artificial intelligence enabled digital banks: A review of expectation confirmation model. *Journal of Enterprise Information Management* 37: 1763–87.

Bond, C. (2024) 20 biggest banks in America – Learn how many branches and ATMs the largest banks have. Retrieved January 16, 2025, from https://www.usnews.com/banking/articles/biggest-banks-in-america

Capital One (2025) Ask Eno: 10 things your Capital One assistant can do for you. Retrieved April 28, 2025, from https://www.capitalone.com/learn-grow/money-management/things-to-ask-eno/

Chalyi, O. (2024). An evaluation of general-purpose AI chatbots: a comprehensive comparative analysis. *InfoScience Trends*, 1(1), 52-66.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, *20*(1), 37-46.

Consumer Financial Protection Bureau. (2023). Chatbots in Consumer Finance. CFPB.

Graham, G., Nisar, T. M., Prabhakar, G., Meriton, R., & Malik, S. (2025). Chatbots in customer service within banking and finance: Do chatbots herald the start of an AI revolution in the corporate world? *Computers in Human Behavior*, 165, 108570.

Gyau, E. B., Appiah, M., Gyamfi, B. A., Achie, T., & Naeem, M. A. (2024). Transforming banking: Examining the role of AI technology innovation in boosting banks financial performance. *International Review of Financial Analysis*, 96, 103700.

Lee, G. U., Hong, D. Y., Kim, S. Y., Kim, J. W., Lee, Y. H., Park, S. O., & Lee, K. R. (2024). Comparison of the problem-solving performance of ChatGPT-3.5, ChatGPT-4, Bing Chat, and Bard for the Korean emergency medicine board examination question bank. *Medicine*, 103(9), e37325.

Johnson, C. and Martin, A. (2025) An AI-powered chatbot for customers helped Goldman Sachs' Marcus realize 'massive savings,' according to a top exec. Here's how. Retrieved April 28, 2025, from https://www.businessinsider.com/goldman-sachs-marcus-using-ai-powered-chatbots-manage-growth-2022-3

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159-174.

Marcus by Goldman Sachs (2025) Here Come the AI Assistants. Retrieved April 28, 2025, from https://www.marcus.com/us/en/resources/heard-at-gs/here-come-the-ai-assistants

Meet TD Clari (2025), Get instant answers to your TD banking questions. Retrieved April 28, 2025, from https://www.td.com/ca/en/personal-banking/solutions/ways-to-bank/chatbot

MX Technologies Inc (2025), Largest U.S. Banks by Asset Size. Retrieved April 28, 2025, from https://www.mx.com/blog/biggest-banks-by-asset-size-united-states/

Qureshi, O. A., Wilkins, S., & Iqbal, H. (2024). Chatbots: Can They Satisfy Customers in the Banking Sector? In BUiD Doctoral Research Conference 2023 (Lecture Notes in Civil Engineering, vol. 473, pp. 360–376). First online March 29, 2024. https://doi.org/10.1007/978-3-031-56121-4_34

Shahriar, S., & Hayawi, K. (2023). Let's have a chat! A Conversation with ChatGPT: Technology, Applications, and Limitations. *arXiv preprint arXiv:2302.13817*.

Truist (2025), Truist launches Truist Assist. Retrieved April 28, 2025, from https://media.truist.com/2022-09-15-Truist-launches-Truist-Assist

Wube, H. D., Esubalew, S. Z., Weldesellasie, F. F., & Debelee, T. G. (2022). Text-Based Chatbot in Financial Sector: A Systematic Literature Review. *Data Science in Finance and Economics*, 2(3), 232–259.

Wu, Haoqi. (2024). Impact of Chatbot Service on Bank Performance Based on a Case Study of IBM Corporation. *Highlights in Business, Economics and Management*, 40, 858-866.

Yatawara, K., Sampath, T., Kalupahana, P. L., Rathnayake, S., Jayasuriya, N., & Rathnayake, N. (2025). A Systematic Review on Consumer Adoption of AI-driven Chatbots. *Vision: Journal of Business Perspective*. https://doi.org/10.1177/09722629251332349

Zhao, Q., Tsai, P.-H., & Wang, J.-L. (2019). Improving financial service innovation strategies for enhancing china's banking industry competitive advantage during the fintech revolution: A Hybrid MCDM model. *Sustainability*, 11(5), 1419.

Zúñiga Salazar, G., Zúñiga, D., Vindel, C. L., & Pérez-Rodríguez, J. (2023). Efficacy of AI chats to determine an emergency: A comparison between OpenAI's ChatGPT, Google Bard, and Microsoft Bing AI Chat. *Cureus*, 15(9), e45473.