EdinBurg Peer Reviewed Journals and Books Publishers
Journal of Information and Technology
Vol. 5||Issue 6||pp 44-55||August||2025
Email: info@edinburgjournals.org||ISSN: 3080-9576

Proposing a Unified Framework for Evaluating Chatbot Efficiency in Banking and Insurance Industries

Ian Pedro Howard Jr[1*], Djuma Sumbiri[2], KN Jonathan[3]
[1,2,3] Department of Information Technology, Faculty of Computing and Information Sciences
University of Lay Adventists of Kigali, Rwanda
Corresponding Author Emails: ianhowardjr@gmail.com;
sumbirdj@gmail.com,phialn1@gmail.com

Accepted: 08 June 2025 || Published: 20 August 2025

Abstract

The integration of artificial intelligence (AI) technologies, particularly chatbots, has transformed customer service operations across various sectors, with banking and insurance standing out due to their high dependence on data accuracy, regulatory compliance, and customer engagement. In these sectors, chatbots are not merely add-ons but integral tools for streamlining service delivery, enhancing user satisfaction, reducing operational costs, and maintaining 24/7 availability. Despite their increasing adoption, the effectiveness of chatbots in these industries is frequently evaluated using fragmented and sector-specific methodologies, leading to inconsistencies in performance assessment and implementation strategies. The absence of a unified evaluation framework undermines the ability of stakeholders to compare chatbot deployments, adopt best practices, and optimize performance in line with strategic business goals. This paper proposes a comprehensive, unified evaluation framework designed specifically for banking and insurance chatbot applications. It incorporates both general performance indicators as response accuracy, response time, natural language processing quality, and user satisfaction industry-specific metrics, including transaction success rate, fraud detection efficiency, regulatory compliance adherence, policy recommendation accuracy, and claims processing efficiency. The framework was developed through a systematic literature review, comparative industry analysis, and synthesis of performance criteria identified in academic and professional research. Although conceptual at this stage, it offers a scalable and adaptable model suitable for real-world applications. Future research will be necessary to empirically validate this model and refine it through iterative field testing and expert feedback. By offering a standardized, multidimensional tool for evaluating chatbot performance, this paper contributes to the broader discourse on AI deployment in critical service industries and supports financial institutions in harnessing the full potential of chatbot technology.

Keywords: *Chatbots, banking, insurance, AI in customer service, chatbot efficiency, unified evaluation framework, financial services*

How to Cite: Howard, I. P., Sumbiri, D., & Jonathan, K. N. (2025). Proposing a Unified Framework for Evaluating Chatbot Efficiency in Banking and Insurance Industries. *Journal of Information and Technology, 5*(6), 44-55.

1. Introduction

Over the past decade, the emergence of AI-powered chatbots has revolutionized the customer service paradigm. These tools are now indispensable in various high-touch, data-driven industries, notably banking and insurance. With the increasing demand for instant responses, personalized services, and secure interactions, chatbots provide a scalable solution to meet these requirements. They perform functions ranging from answering frequently asked questions and processing transactions to delivering personalized financial advice and managing insurance claims.

In banking, chatbots enable real-time engagement by automating routine interactions like balance inquiries, transaction history checks, and even facilitating financial planning (Eustaquio-Jiménez et al., 2024; Wu, 2024). They also play a critical role in enhancing fraud prevention through real-time transaction monitoring and customer verification (Bokolo & Daramola, 2024).

Similarly, in insurance, chatbots are being used to simplify and automate complex procedures such as claims submission, underwriting assessments, and policy renewals (Chen, 2025; Eustaquio-Jiménez et al., 2024). However, the benefits of chatbot deployment can only be fully realized if their performance is rigorously evaluated against clearly defined criteria. Currently, most institutions use bespoke methods to assess chatbot success, which are often tailored to specific use cases or technological architectures (Zainol et al., 2023). This piecemeal approach makes it difficult to benchmark performance across implementations, stifles innovation, and can obscure deficiencies that affect customer experience and operational outcomes.

This paper argues for a unified evaluation framework tailored to the banking and insurance sectors. It identifies key performance metrics that are applicable across both industries while also acknowledging and incorporating industry-specific requirements. The proposed framework aims to standardize chatbot evaluation, facilitate best practice sharing, and drive technological improvements aligned with organizational goals.

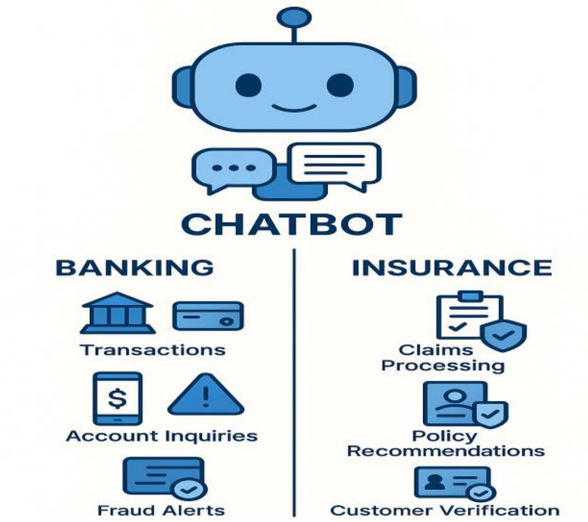

Figure 1: Unified framework for Banking and Insurance Industries

EdinBurg Peer Reviewed Journals and Books Publishers
Journal of Information and Technology
Vol. 5||Issue 6||pp 44-55||August||2025
Email: info@edinburgjournals.org||ISSN: 3080-9576

2. Literature Review

2.1 Chatbots in Banking

The banking sector has been among the earliest adopters of chatbot technology due to the high volume of routine customer inquiries and the critical need for real-time engagement. According to Eustaquio-Jiménez et al. (2024), banks employ chatbots to manage common queries about account balances, fund transfers, credit card details, and loan applications. This allows human agents to focus on complex queries that require deeper analysis or empathy.

Banks also leverage chatbots for customer onboarding, helping users through the process of opening accounts or applying for loans with minimal human intervention. More advanced implementations feature conversational AI that provides tailored financial advice based on user behavior and financial goals (Wu, 2024; Munira et al., 2025).

However, the high sensitivity of financial data poses unique challenges for chatbot deployment in banking. According to Edwards (Maroengsit et al., 2019), any mishandling of user data, errors in financial transactions, or breaches in communication security can lead to significant legal and reputational damage. Consequently, banking chatbots must adhere to stringent regulatory frameworks.

Moreover, building user trust is vital. Unlike e-commerce or retail, where errors may lead to minor inconveniences, failures in banking chatbot interactions can result in significant financial loss or misinformation. Hence, chatbots must be accurate, transparent, and secure (Zainol et al., 2023).

2.2 Chatbots in Insurance

The insurance industry has also embraced chatbot technology, albeit with a slightly different set of priorities. The focus here is often on enhancing accessibility, simplifying policy management, and expediting the claims process. Alkhelb and Alshagrawi (2025) emphasize the potential of chatbots to automate claims filing, enabling users to submit details and upload documents without navigating cumbersome forms or waiting in call center queues.

Chatbots also assist in guiding customers through the selection of suitable insurance policies by asking a series of intelligent questions and recommending products based on user inputs. This improves both transparency and customer satisfaction. de Andrés-Sánchez & Gené-Albesa (2024) point out that insurance chatbots serve a dual role as both educators and facilitators, demystifying complex insurance jargon and assisting users in making informed decisions.

Yet, challenges abound. Insurance claims are complex, context-sensitive, and often emotionally charged. Chatbots must be capable of interpreting nuanced human input and responding with both accuracy and empathy (Wu, 2024). Furthermore, detecting fraudulent claims is a major concern in the insurance domain. Wilson and Hunter (2025) note that some insurers have begun integrating machine learning into chatbot backends to flag anomalies indicative of potential fraud.

Privacy and compliance are also critical. In many jurisdictions, insurance providers must comply with data protection laws similar to or even more stringent than those in banking. Chatbots must be designed to store, process, and transmit personal information in a compliant and secure manner (Chen, 2025).

EdinBurg Peer Reviewed Journals and Books Publishers
Journal of Information and Technology
Vol. 5||Issue 6||pp 44-55||August||2025
Email: info@edinburgjournals.org||ISSN: 3080-9576

The Evaluation Gap

While individual studies have explored performance metrics for chatbots in either banking or insurance, few have attempted to create a consolidated framework that encompasses both. This siloed approach leads to redundancy, hinders cross-sector benchmarking, and results in missed opportunities to leverage insights from parallel implementations (Maroengsit et al., 2019). The need for a unified, adaptable framework is clear.

3. Methodology

The methodology employed in this study is based on qualitative research principles, primarily drawing on secondary data sources such as scholarly articles, industry reports, white papers, and documented case studies. The research unfolds in three stages:

1. **Systematic Literature Review** – A comprehensive literature review was conducted using databases such as IEEE Xplore, ScienceDirect, SpringerLink, and Google Scholar. Selection criteria required that articles address performance metrics, implementation challenges, and use cases relevant to the two target industries (Maroengsit et al., 2019; Munira et al., 2025).
2. **Comparative Sector Analysis** – The collected data were synthesized to identify overlapping and unique evaluation criteria across banking and insurance.
3. **Framework Development** – The insights will be structured into a cohesive framework composed of general and sector-specific performance metrics, mapped onto a proposed evaluation model (Zainol et al., 2023).

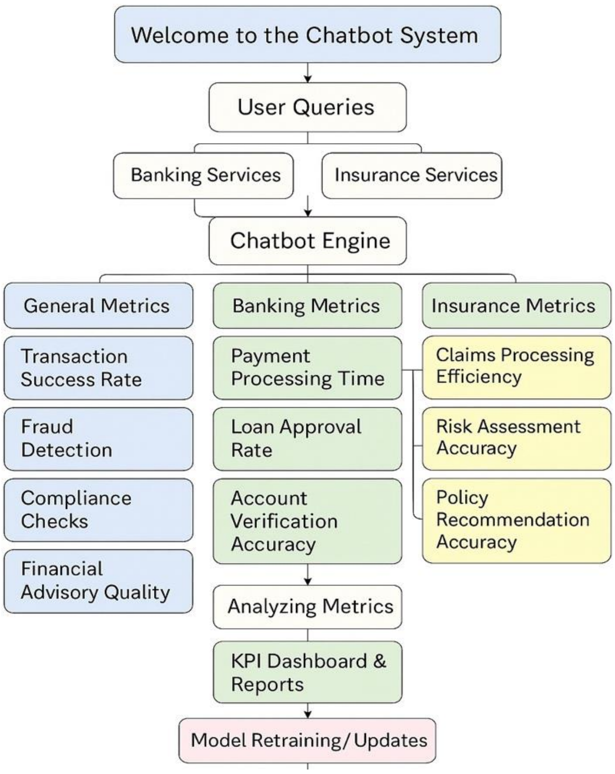

Figure 2: **Flowchart of Unified Chatbot Evaluation Framework for Banking and Insurance**

EdinBurg Peer Reviewed Journals and Books Publishers
Journal of Information and Technology
Vol. 5||Issue 6||pp 44-55||August||2025
Email: info@edinburgjournals.org||ISSN: 3080-9576

To support implementation, the system-level architecture of this framework is depicted in Figure 3 below.

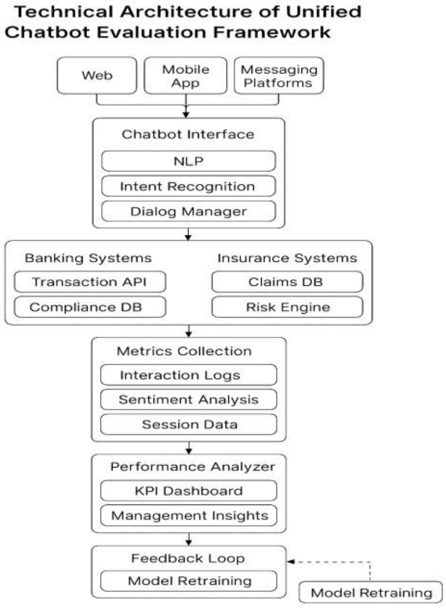

Figure 3: Technical Architecture of Unified Chatbot Evaluation Framework For Banking and Insurance

4. Proposed Evaluation Framework

To establish a robust and standardized methodology for assessing chatbot efficiency within the banking and insurance industries, this paper introduces a dual-structured evaluation framework comprising both general performance metrics and industry-specific indicators. These criteria aim to capture not only the functional performance of chatbots but also their ability to satisfy users, comply with regulations, and adapt to evolving service needs. This section expands each metric in detail to demonstrate its significance, method of evaluation, and potential implications.

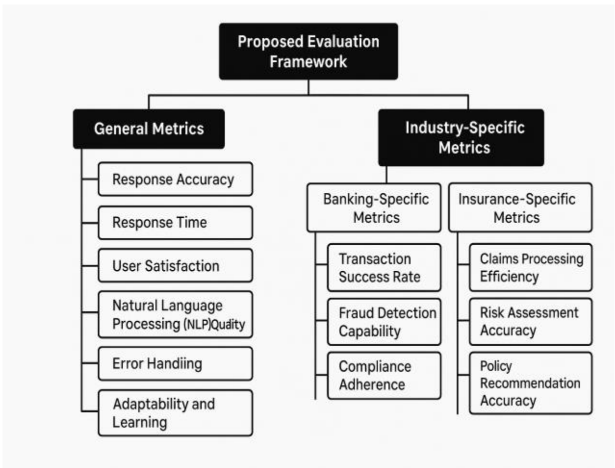

Figure 4: Proposed Evaluation Framework

4.1. General Performance Metrics

1. Response Accuracy

Response accuracy is arguably the most critical metric in any chatbot performance framework. It refers to the chatbot's ability to provide correct, contextually appropriate, and relevant responses to user inquiries. In the financial services sector, where precision is paramount, inaccurate information can lead to serious consequences, including financial loss, customer dissatisfaction, and regulatory breaches (Eustaquio-Jiménez et al., 2024; Bokolo & Daramola, 2024). For example, if a banking chatbot provides the wrong account balance or an insurance chatbot misinforms a user about coverage limitations, the fallout could be significant. Evaluating this metric typically involves analyzing historical chat logs, auditing automated conversations for correctness, and scoring them against a verified dataset of expected outputs. Additionally, modern NLP techniques can help identify misclassifications in user intent, providing diagnostic feedback for model improvement.

2. Response Time

Response time measures how quickly a chatbot responds to user inputs. While this might seem like a technical detail, it has profound implications for user satisfaction and perceived efficiency (Maroengsit et al., 2019). In the age of instant communication, users expect real-time interaction, especially when dealing with urgent issues such as suspicious transactions or emergency claims. A delay of even a few seconds can disrupt user trust and negatively affect satisfaction scores. Response time is typically tracked in milliseconds and can be broken down into server processing time and front-end delivery latency. Reducing response time often involves streamlining backend processes, optimizing code execution, and deploying faster infrastructure such as low-latency APIs or edge computing solutions.

3. User Satisfaction

User satisfaction is a holistic measure that reflects how end-users feel about their interactions with the chatbot. This includes perceptions of ease of use, reliability, clarity, empathy, and overall effectiveness in solving their problems (Wu, 2024; Zainol et al., 2023). In the banking context, satisfaction may stem from seamless transactions and responsive security protocols, while in insurance, it may depend on how effectively the chatbot guides users through claims or policy renewals. Methods for evaluating this metric include post-interaction surveys, Net Promoter Scores (NPS), sentiment analysis of free-text feedback, and user retention metrics. Satisfaction is not only a reflection of chatbot quality but also a predictive indicator of brand loyalty and customer lifetime value.

4. Natural Language Processing (NLP) Quality

The quality of a chatbot's natural language processing (NLP) capability directly influences its ability to interpret user intent and provide meaningful responses. High NLP quality means the chatbot can handle slang, abbreviations, complex grammar, and multi-turn dialogues without breaking context (Maroengsit et al., 2019; Zainol et al., 2023). Poor NLP quality leads to user frustration, repeated queries, and eventual disengagement. This metric is typically measured using machine learning performance indicators such as precision, recall, F1 score, BLEU score, and perplexity. These metrics can be validated against test corpora and real-world interaction datasets. Furthermore, qualitative evaluationslike conversation replay auditscan help determine whether the bot maintains conversational coherence and emotional appropriateness in sensitive contexts such as fraud alerts or denied claims.

5. Error Handling

Effective error handling is crucial in determining whether a chatbot can maintain user engagement even when it fails to understand a query or fulfill a request. Robust error-handling mechanisms ensure that failures don't escalate into frustration. Instead, they become opportunities to steer the user back on track through clarifications or escalations to human support (Bokolo & Daramola, 2024; Zainol et al., 2023). For instance, when an insurance chatbot fails to recognize an unusual claim type, it should respond with an empathetic prompt and offer to connect the user to an agent, rather than returning a generic "I don't understand" message. Evaluations of this metric often involve assessing fallback flow effectiveness, dropout rates after failure prompts, and qualitative user sentiment following error messages.

6. Adaptability and Learning

This metric assesses the chatbot's ability to evolve by learning from user interactions, integrating new content, and improving intent recognition models. Chatbots that incorporate machine learning and continual improvement pipelines demonstrate long-term viability and operational excellence (Bokolo & Daramola, 2024; Vuković et al., 2025). Adaptability also implies the chatbot can recognize when its existing knowledge is outdatedfor example, identifying when financial policies have changedand seek updates from its training pipeline or connected databases. Evaluation criteria include performance trend analysis, decline in repeated failures, user feedback integration rates, and retraining frequency. Adaptive systems are especially important in dynamic sectors like banking and insurance, where product lines, compliance requirements, and user behavior evolve regularly.

4.2. Banking-Specific Metrics

1. Transaction Success Rate

This metric evaluates the chatbot's effectiveness in completing financial transactions such as transfers, bill payments, and loan inquiries. High transaction success rates indicate that the chatbot is both well-integrated with backend systems and capable of correctly interpreting user intent (Maroengsit et al., 2019). In contrast, a low success rate may point to system limitations, NLP inadequacies, or workflow design flaws. Success is generally measured as the percentage of transaction intents that are completed without human assistance. Tracking this metric allows banks to identify friction points and prioritize improvements in high-volume services.

2. Fraud Detection Capability

In an increasingly digital financial landscape, fraud detection is a mission-critical function. While most fraud detection logic resides in backend systems, chatbots serve as an important user-facing layer that can collect contextual clues or behavioral patterns to flag suspicious activities (Vuković et al., 2025). For example, if a user attempts to transfer a large sum to an unfamiliar account while traveling abroad, the chatbot might ask for additional verification or trigger a security alert. Evaluating this metric includes tracking the number of successful fraud preventions initiated through chatbot interactions, false positive/negative rates, and escalation timelines. As AI-driven financial services become more prevalent, incorporating proactive fraud detection capabilities within chatbots adds a layer of user protection and institutional accountability (Maroengsit et al., 2019).

3. Compliance Adherence

Given the regulatory intensity of the banking sector, chatbot compliance with local and international laws is non-negotiable. This includes the General Data Protection Regulation (GDPR), the Payment Card Industry Data Security Standard (PCI DSS), and Know Your Customer (KYC) protocols. The chatbot must be capable of obtaining and recording user consent, presenting legally mandated disclosures, and executing data deletion requests (Maroengsit et al., 2019; Zainol et al., 2023). Additionally, compliance adherence entails that the chatbot never performs unauthorized actions and logs all transactional activities in audit-friendly formats. Evaluation involves legal audit simulation, compliance checklist scoring, and real-time alerts for policy violations.

4. Financial Advisory Quality

Advanced banking chatbots often provide financial planning tools that offer budget advice, suggest savings strategies, or recommend investment opportunities (Bokolo & Daramola, 2024; de Andrés-Sánchez & Gené-Albesa, 2024). This metric evaluates how well these recommendations align with user profiles and goals. For example, suggesting high-risk equity investments to a conservative user is a sign of poor advisory logic. Key performance indicators include conversion rates, acceptance of financial advice, feedback scores, and long-term return-on-investment (ROI) tracking. High-quality financial advice not only drives user engagement but also positions the institution as a trusted partner in personal finance management.

4.3. Insurance-Specific Metrics

1. Claims Processing Efficiency

This metric assesses the chatbot's ability to guide users through the entire claims process—from initial reporting to document submission and follow-up inquiries (Chen, 2025; Eustaquio-Jiménez et al., 2024). Efficient claims processing reduces administrative costs, shortens resolution timelines, and enhances customer satisfaction. Evaluations typically involve tracking average handling times, claim initiation-to-closure duration, and user drop-off rates. Additionally, this metric may incorporate backend integration success rates and error-free submission percentages.

2. Risk Assessment Accuracy

Insurance products are priced and approved based on risk evaluation. Chatbots increasingly assist in this process by asking users pre-underwriting questions and compiling basic risk profiles (Chen, 2025). Accuracy in this function ensures that policies are appropriately priced and that users are neither under- nor over-insured. This metric is evaluated by comparing chatbot-generated risk assessments with those produced by human underwriters and measuring alignment scores. Misalignment can lead to financial loss or legal exposure, especially in high-risk segments like health or travel insurance.

3. Policy Recommendation Accuracy

This refers to the chatbot's effectiveness in suggesting appropriate insurance products based on user input and historical interaction data (Chen, 2025). A well-performing chatbot should recommend policies that users are likely to accept and that match their stated needs and demographic profile. Evaluation indicators include policy purchase conversion rates, session duration, and bounce rates following recommendation interactions. Feedback loops, such as post-purchase surveys, can help validate the perceived value of the chatbot's suggestions.

4. Customer Sentiment During Claims

The claims process is a high-emotion touchpoint where user perception is fragile. This metric analyzes how the chatbot manages user sentiment throughout the claims lifecycle (Sodré & Duarte, 2023). It involves sentiment analysis of text inputs, escalation tracking, and post-interaction surveys. Chatbots must balance empathy with efficiency, providing timely updates and reassurances without appearing robotic or indifferent. Poor sentiment handling can escalate frustration and result in complaints or customer churn, while excellent handling can turn a negative event into a brand-strengthening experience.

5. **Evaluation Methodology**

To systematically assess chatbot performance in banking and insurance, the following five-stage methodology is proposed. Each phase contributes unique insights, ensuring a multidimensional understanding of chatbot efficiency:

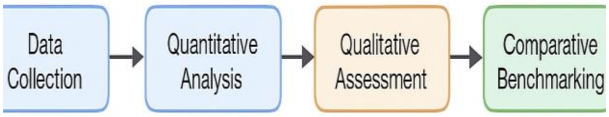

Figure 5*:* Evaluation Methodology for Assessing Chatbot Efficiency

1. **Data Collection:** The foundation of effective evaluation is high-quality, comprehensive data. This includes chatbot interaction logs, session durations, user queries, resolution outcomes, transaction attempts, error reports, and post-chat survey responses. Collected data should also account for escalations to human agents, user sentiment scores, and compliance flags. Proper anonymization and ethical data handling are essential, particularly when working with sensitive financial or personal information. Advanced logging frameworks can help automate this process and ensure consistency across chatbot deployments.
2. **Quantitative Analysis:** This phase involves statistically analyzing numerical indicators such as response accuracy, transaction success rates, average handling time, and conversation dropout rates. Tools such as dashboards, heat maps, and trend graphs are commonly used to visualize performance. These metrics provide objective, repeatable assessments and help identify performance trends or anomalies. In financial sectors, quantitative indicators also enable early detection of risk areas, such as unusually high failure rates during peak transaction times.
3. **Qualitative Assessment:** While quantitative data tells what happened, qualitative data explains why it happened. This step involves analyzing user feedback, open-ended survey responses, and expert reviews to assess user perception, emotional satisfaction, and conversation quality (Wu, 2024; Munira et al., 2025). Methods include sentiment

EdinBurg Peer Reviewed Journals and Books Publishers
Journal of Information and Technology
Vol. 5||Issue 6||pp 44-55||August||2025
Email: info@edinburgjournals.org||ISSN: 3080-9576

analysis, scenario walkthroughs, and usability testing. This phase is particularly crucial in emotionally sensitive contexts, such as insurance claims or fraud inquiries, where empathy and clarity directly impact user trust.

4. **Comparative Benchmarking**: Institutions should compare their chatbot's performance against internal goals, historical trends, and industry standards. Benchmarking enables organizations to identify strengths and weaknesses relative to competitors or peers. It also provides a contextual baseline for evaluating new chatbot features or retraining impacts.
5. **Continuous Optimization**: Evaluation should not be a one-time process. Chatbots must evolve through feedback loops that incorporate new training data, updated intents, and refined NLP models. Regular performance reviews should lead to iterative improvements in conversational design, compliance alignment, and system integration. This ensures that the chatbot remains relevant, effective, and future-proof.

6. Discussion

The widespread adoption of AI-powered chatbots in financial services has made performance evaluation not just desirable but essential. With increasingly complex user needs and heightened regulatory scrutiny, the ability to measure and refine chatbot performance becomes a core component of digital transformation strategies. The proposed unified evaluation framework brings structure and clarity to what has been a fragmented field. It allows financial institutions to apply a consistent lens to chatbot efficiency, enabling benchmarking across departments, use cases, and even industries.

One of the key contributions of the framework is its dual-layer design, separating general performance indicators from domain-specific metrics. This modularity ensures that core functionalities like response accuracy and NLP quality are not evaluated in isolation but are complemented by industry-relevant considerations such as fraud detection or claims sentiment (Zainol et al., 2023; Vuković et al., 2025). By maintaining this balance, the framework accommodates both technical performance and user-centricity.

Moreover, the framework promotes continuous learning. Metrics such as adaptability encourage institutions to develop chatbots that are not static tools but dynamic systems capable of evolving alongside user behavior, product changes, and legal requirements (Bokolo & Daramola, 2024; Vuković et al., 2025). This aligns with emerging trends in AI ethics, where accountability, transparency, and improvement are seen as foundational pillars of responsible deployment.

Ultimately, this framework is more than a checklist—it is a strategic blueprint. It empowers stakeholders to make data-driven decisions, foster innovation, and deliver superior customer experiences. As chatbot technology matures and penetrates deeper into financial workflows, the ability to evaluate and refine its performance will be a defining factor in determining organizational success in digital customer engagement.

7. Future Research and Validation

Although the proposed framework presents a robust foundation for evaluating chatbot efficiency in the banking and insurance sectors, it is currently conceptual and requires empirical validation through real-world application. Future research should focus on implementing this framework across a diverse range of financial institutions to test its applicability, adaptability, and overall impact on chatbot performance outcomes.

One immediate avenue for future research is the development of standardized benchmarks for each performance metric. These benchmarks could be industry-wide averages or best-in-class scores derived from leading institutions. By establishing these reference points, organizations will be better equipped to measure relative performance and set realistic targets for chatbot improvements.

In addition, longitudinal studies should be conducted to observe how chatbot performance evolves following the implementation of the framework. These studies could measure improvements in transaction success rates, fraud detection capabilities, or customer satisfaction, offering evidence for the framework's efficacy and highlighting any areas needing refinement.

Another critical area for future exploration is cross-platform and cross-channel integration. Chatbots are increasingly being deployed across mobile apps, websites, and messaging platforms. Research should investigate how the framework accommodates multichannel behavior and whether performance varies significantly depending on the delivery medium.

User segmentation represents another layer of analysis. Different user groups—based on age, income, digital literacy, or geography—may interact differently with chatbots. Tailoring evaluation criteria to specific demographics could uncover insights that improve chatbot personalization and equity in service delivery.

Lastly, collaborative validation with regulators and industry bodies will enhance the framework's credibility and compliance relevance. Incorporating regulatory feedback ensures that the framework remains aligned with evolving legal and ethical standards, such as data protection laws and AI accountability principles.

These research directions will help convert the proposed evaluation framework from a theoretical model into an industry-standard tool for chatbot optimization in financial services.

8. Conclusion

In conclusion, the unified evaluation framework presented in this paper addresses a critical gap in the deployment and optimization of chatbots within the banking and insurance industries. By integrating both general and sector-specific performance metrics, the framework enables a structured, scalable, and context-aware approach to performance analysis. It ensures that chatbot systems are not only functionally competent but also aligned with user needs, institutional objectives, and regulatory mandates. Furthermore, by emphasizing adaptability, error handling, and sentiment responsiveness, the framework advances the development of emotionally intelligent and ethically sound AI systems. As chatbots become a strategic component of customer service architecture, the ability to evaluate their performance rigorously will determine whether organizations can transform automation into meaningful, user-centered engagement. This framework thus provides the foundation for sustained innovation, compliance, and service excellence in the digital era (Maroengsit et al., 2019; Zainol et al., 2023; Wu, 2024).

EdinBurg Peer Reviewed Journals and Books Publishers
Journal of Information and Technology
Vol. 5||Issue 6||pp 44-55||August||2025
Email: info@edinburgjournals.org||ISSN: 3080-9576

References

[1] Bokolo, Z., & Daramola, O. (2024). Elicitation of security threats and vulnerabilities in insurance chatbots using STRIDE. *Scientific Reports*, *14*(1), Article 10482. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11297133/

[2] Chen, Z. (2025). Revolutionizing finance with conversational AI: A focus on ChatGPT implementation and challenges. *Humanities and Social Sciences Communications*, *12*(1), Article 152. https://www.nature.com/articles/s41599-025-02761-2

[3] de Andrés-Sánchez, J., & Gené-Albesa, J. (2024). Not with the bot! The relevance of trust to explain the acceptance of chatbots by insurance customers. *Humanities and Social Sciences Communications*, *11*(1), Article 104. https://www.nature.com/articles/s41599-024-02515-z

[4] Eustaquio-Jiménez, R., Durand-Azurza, M., Gamboa-Cruzado, J., León Morales, M., Pajares Ruiz, N., & López de Montoya, R. (2024). Chatbots for customer service in financial entities—A comprehensive systematic review. *Journal of Infrastructure, Policy and Development*, *8*(16), Article 10122. https://systems.enpress-publisher.com/index.php/jipd/article/view/10122

[5] Maroengsit, W., Piyakulpinyo, T., Phonyiam, K., & Pongnumkul, S. (2019). A survey on evaluation methods for chatbots. In *Proceedings of the 2019 7th International Conference on Information and Education Technology (ICIET)* (pp. 7–11). ACM. https://dl.acm.org/doi/10.1145/3323771.3323824

[6] Munira, M. S. K., Juthi, S., & Begum, A. (2025). Artificial intelligence in financial customer relationship management: A systematic review of AI-driven strategies in banking and FinTech. *American Journal of Advanced Technology and Engineering Solutions*, *1*(1), 1–10. https://ajates-scholarly.com/index.php/ajates/article/view/3

[7] Sodré, W. S. M., & Duarte, J. C. (2023). Chatbot optimization using sentiment analysis and timeline navigation. *Revista de Informática Teórica e Aplicada*, *30*(1), 73–93. https://seer.ufrgs.br/rita/article/view/125825

[8] Vuković, D. B., Dekpo-Adza, S., Alghushairy, A., & Morales, D. (2025). AI integration in financial services: A systematic review of trends and regulatory challenges. *Humanities and Social Sciences Communications*, *12*(1), Article 128. https://www.nature.com/articles/s41599-025-02751-4

[9] Wu, H. (2024). Impact of chatbot service on bank performance based on a case study of IBM Corporation. *Highlights in Business, Economics and Management*, *40*, 276–281. https://drpress.org/ojs/index.php/HBEM/article/view/24732

[10] Zainol, S., Shamsudin, M. F., Hassan, S., & Mohd Noor, N. A. (2023). Understanding customer satisfaction with chatbot service and system quality in banking services. *Journal of Information Technology Management*, *15*(SI), 93–110. https://www.researchgate.net/publication/373841113