International Journal of Bank Marketing

From conversation to insight: competitive learning and generative AI for customer experience extraction in the financial sector

Received 21 October 2025
Revised 29 January 2026
11 April 2026
17 April 2026
Accepted 17 April 2026

José-de-Jesús Rocha-Salazar
Tecnologico de Monterrey, EGADE Business School, Monterrey, Mexico and Facultad de Ciencias Físico Matemáticas, Universidad Autonoma de Nuevo Leon, San Nicolas de los Garza, Mexico

Jatinder-Jit Singh
Tecnologico de Monterrey, EGADE Business School, Monterrey, Mexico, and

María-Jesús Segovia-Vargas
Faculty of Economics and Business Administration, Universidad Complutense de Madrid, Madrid, Spain

Abstract

Purpose – Customer experience (CX) is a key driver of cost efficiency, revenue growth, and competitive advantage in the banking industry. From a service-dominant logic and relationship marketing perspective, experience emerges through repeated interactions in which value is co-created over time. However, traditional tools such as surveys and Net Promoter Score (NPS) provide only partial insights into value-in-use and relationship value and often fail to capture customers' lived experiences. This study addresses this limitation by proposing a data-driven approach that directly captures customer experience from unstructured conversational data.

Design/methodology/approach – The study combines competitive learning and generative artificial intelligence (GenAI) to analyze 10,062,406 customer–agent conversations collected monthly by a financial institution between July 2023 and July 2025. The proposed system allows customer experience specialists to query a conversational agent in natural language and receive context-rich responses, enabling exploratory customer experience analysis.

Findings – The results show that a conversational agent created through competitive learning and GenAI yields richer, more actionable insights into customer experiences than conventional methods. The framework demonstrates strong validity based on human evaluation and AI benchmarking against Gemini and DeepSeek (human scores >4.4; similarity >0.9 with Gemini; similarity >0.85 with DeepSeek). Compared with BERTopic and Top2Vec, it achieves higher computational efficiency, lower costs, and scalable, interpretable outputs for managerial use.

Originality/value – This study is among the first to apply competitive learning and GenAI to customer experience analytics in banking, offering a scalable pathway to improve service performance and competitive positioning.

Keywords Banking, Customer experience, Artificial intelligence, Competitive learning, Natural language, Clustering

Paper type Research article

International Journal of Bank Marketing
Emerald Publishing Limited
e-ISSN: 1758-5937
p-ISSN: 0265-2323
DOI 10.1108/IJBM-10-2025-0815

1. Introduction

Customer experience (CX) has become a central source of competitive differentiation in banking because it shapes customer satisfaction, trust, loyalty, and longer-term firm performance. Prior studies, such as Komulainen and Saraniemi (2019), Buhler *et al*. (2024), and Hoang (2024), indicate that customer evaluations of banking relationships are influenced by multiple dimensions, including customer centricity in mobile banking and the design of technology-mediated service encounters. These dimensions, together with broader loyalty drivers, have important consequences for downstream behavioral outcomes.

As financial products become increasingly standardized and digitally accessible, competitive advantage has shifted away from product features towards the quality of service interactions through which value is created and delivered. Prior research consistently demonstrates that positive customer experiences are associated with higher retention, increased cross-selling, and improved profitability, whereas negative service encounters accelerate customer attrition and undermine institutional trust (Kumar *et al*., 2022; Manyanga *et al*., 2022; Peruchini *et al*., 2024). Moreover, industry evidence further indicates that approximately 88% of customers consider CX to be as important as, or more critical than, product offerings when selecting a bank, and that poor service interactions account for nearly 20% of customer attrition (10x Banking, 2023; FICO, 2025).

Despite its strategic relevance, CX measurement in banking remains largely grounded in structured, episodic instruments such as surveys, Likert-scale evaluations, and Net Promoter Score (NPS) metrics. These tools were initially designed to capture post hoc assessments of service outcomes rather than the dynamic, interactional processes through which service experiences unfold (Chatterjee, 2019). As a result, they provide only a partial representation of customer experience and offer limited insight into how customer needs, emotions, and expectations emerge and evolve during service encounters. Moreover, survey-based approaches are subject to self-selection bias and typically reflect the perceptions of only a small fraction of customers, leaving substantial experiential information unobserved. These limitations have motivated growing interest in data-driven and AI-enabled approaches capable of capturing customer experience at scale and in real time.

Against this backdrop, the emergence of Industry 4.0 technologies has enabled banks to enhance traditional CX analysis through more sophisticated data science and artificial intelligence solutions (Adeniran *et al*., 2024). Hentzen *et al*. (2022), Mogaji and Nguyen (2021), and An (2026) suggest that artificial intelligence in customer-facing financial services has moved beyond a purely conceptual promise and has become a matter of managerial relevance and service design. Banks are increasingly assessing how AI can support marketing, customer interactions, and service quality in financial services contexts. Evidence of this can be found in machine learning models that have provided more accurate approaches to sentiment analysis, churn prediction, identification of statistically significant factors affecting satisfaction, and assessment of perceived service quality. Although more robust, these models still rely on structured or unstructured data as inputs and produce limited, highly structured outputs.

To change the structured nature of analytical outputs, conversational agents, primarily chatbots, have emerged to generate more natural language responses and interactions. However, the literature indicates that these chatbots have focused on managing and improving customer interactions rather than extracting customer experience (CX) insights, and they continue to exhibit performance and reliability limitations (Bhatnagr and Rajesh, 2025).

The preceding discussion reveals a conceptual misalignment between current CX research and practice. Customer experience in banking is increasingly produced through ongoing, multi-turn interactions across multiple channels. However, it is still analyzed using static, outcome-oriented measurement frameworks. Moreover, most existing conversational agent solutions are not explicitly designed to extract customer experience insights. Instead, they focus on optimizing operational tasks and satisfaction outcomes, thereby overlooking the interactional processes through which customer experience is formed (Komulainen and

Makkonen, 2018; Hamdan, 2025). Relatedly, recent work on generative AI in banking has mainly examined questions of organizational integration, reliability, responsiveness, and regulatory compliance; while still leaving unresolved the question of how these technologies can be used to derive analytically grounded customer-experience insights from large-scale customer–service conversations (Moharrak and Mogaji, 2025).

To address this gap, the present study proposes a framework that integrates competitive learning with generative artificial intelligence (GenAI) to extract customer experience insights from customer–agent conversations. From a marketing and service management perspective, customer–agent conversations constitute a rich empirical source, providing direct linguistic evidence of trust, frustration, uncertainty, and satisfaction as these states are articulated during real service encounters. Consistent with service-dominant logic (S-D logic), value in banking is not embedded in products or service outputs. Still, it is co-created through interactions between customers and service agents through the contact center, particularly as customers seek problem resolution and goal attainment. In this regard, customer–agent conversations represent critical moments of value co-creation in which customers integrate institutional resources with their own expectations, competencies, and emotional states to construct their overall experience (Vargo and Lusch, 2016, 2017).

By identifying recurring interactional conversational contexts at scale, the proposed approach operationalizes S-D logic and relationship marketing perspectives by revealing stable patterns through which value co-creation and relationships are reinforced or undermined over time. Unlike customer-facing chatbots, the system is designed as a staff-facing analytical tool that enables customer experience specialists to pose natural language queries and receive context-rich, natural language responses grounded in large-scale conversational evidence.

The framework operates daily on a rolling 31-day window within a global financial institution. For reporting purposes, this article presents only the results obtained on the last day of each month, from July 2023 to July 2025, using a dataset of 10,062,406 client–agent conversations. Validation through human rubric-based evaluation indicates high reliability, with accuracy scores ranging from 4.4 to 5, indicating excellent performance. Complementary AI-based benchmarking further confirms the robustness of the framework, with accuracy levels exceeding 0.85 when compared with DeepSeek and surpassing 0.9 when benchmarked against Gemini throughout the analysis period. Moreover, relative to established topic modeling approaches such as BERTopic and Top2Vec, the proposed methodology demonstrates superior computational efficiency, as evidenced by shorter execution times and lower computational costs.

The remainder of this article is organized as follows. Section 2 reviews the relevant literature on customer experience in the banking sector and outlines the research objectives. Section 3 describes the proposed methodology and data preparation. Section 4 presents the experimental results and subsequent discussion. Finally, Section 5 summarizes the main conclusions, outlines the implications, and discusses the study's limitations along with directions for future research.

2. Literature review and objectives

Customer experience refers to customers' perceptions, emotions, and responses that develop across their interactions with a financial institution throughout the service lifecycle (Le and Nguyen, 2024). Within contemporary service-dominant logic, CX is conceptualized as a processual and relational phenomenon that emerges from continuous service interactions rather than as a static evaluation of discrete products or services. From this perspective, value is not embedded in banking offerings. Still, it is realized in use, as customers integrate institutional resources (e.g. service personnel, digital platforms, products, policies) with their own skills, expectations, and situational contexts (Vargo and Lusch, 2016, 2017). Hijazi (2022) shows, for example, that mobile banking service quality can stimulate customer value co-creation intentions, reinforcing the relevance of examining banking interactions as participatory and relational processes rather than as isolated transactions. Service encounters, therefore, function as critical sites for the realization of value-in-use and the formation of experiential meaning.

S-D Logic further emphasizes that customer experience unfolds within service ecosystems, where repeated interactions are shaped by institutional arrangements, shared norms, and organizational practices that structure how value co-creation occurs over time. In highly regulated sectors such as banking, these institutional conditions play a central role in influencing how customers perceive fairness, transparency, and reliability across service interactions (Vargo and Lusch, 2016). Consequently, CX reflects not only isolated service moments but also customers' cumulative interpretations of how consistently and effectively a financial institution supports their goals and needs across channels and over time.

In parallel, relationship marketing theory provides a complementary lens for understanding the long-term consequences of customer experience. Relationship-oriented research highlights that relational outcomes emerge cumulatively, as repeated service encounters contribute to the development of trust, commitment, satisfaction, and perceived relational value (Palmatier *et al*., 2018). These relational assets strengthen customer loyalty by reducing uncertainty, decreasing churn, and fostering emotional attachment to the service provider. In financial services, where perceived risk and information asymmetry are high, the quality and consistency of ongoing interactions are particularly influential in sustaining durable customer relationships.

The growing strategic emphasis on customer experience (CX) in banking has unfolded in parallel with the sector's broader digital transformation. As CX is increasingly understood as an interactional and relational process that evolves across service encounters, banks have sought analytical approaches capable of capturing customer experiences at scale and in real time. Conventional CX assessment tools, such as surveys and Net Promoter Score (NPS) metrics, are therefore progressively complemented and in some cases supplanted by artificial intelligence (AI)–driven methods that can process unstructured customer interactions and generate actionable insights.

Advances in AI technologies, including natural language processing (NLP), machine learning (ML), and, more recently, generative AI, have expanded organizations' abilities to extract, interpret, and operationalize customer experience signals embedded in large volumes of conversational data.

Reflecting this evolution, academic research on AI applications for CX in banking has developed along three principal streams: sentiment analysis, conversational agents, and generative AI. The following literature review synthesizes recent peer-reviewed studies within these streams, highlighting how AI-based approaches have contributed to the analysis of customer experience, as well as the conceptual and methodological limitations that motivate the present study.

2.1 Sentiment analysis

Text mining and sentiment analysis of unstructured customer feedback are among the earliest and most influential artificial intelligence techniques applied in customer experience (CX) research within the banking sector. Ghadiridehkordi *et al*. (2025) and Çallı (2023) show that unstructured customer text can reveal service-specific satisfaction patterns and service-quality dimensions that are difficult to capture through conventional instruments. Kaur (2021) shows that a sentiment analysis pipeline using NLP and machine learning can effectively classify e-banking customer reviews, though persistent class imbalance limits the accurate detection of negative sentiments. In a related approach, Akter *et al*. (2024) propose a comprehensive, data-driven AI framework for customer experience analytics that integrates sentiment analysis with deep learning. Their findings reveal that applying machine learning to online banking reviews can uncover underlying CX dimensions, including trust, ease of use, and responsiveness.

Expanding on these methodologies, Mittal and Agrawal (2022) conduct regression analysis to examine the effects of multiple service attributes on customer satisfaction ratings, providing explanatory insights into variables with statistically significant impacts on CX. Further methodological diversity is observed in the work of Andrian *et al*. (2022), who develop a binary classification model to predict sentiment in Twitter reviews about digital banking, using demographic features and word-frequency data as input variables. Similarly, Leem and Eum (2021) apply Naive Bayes and other supervised learning techniques for sentiment classification, supplementing these methods with keyword frequency and network analysis to identify thematic structures. Their approach involves determining sentiments in mobile banking application reviews, evaluating classifier accuracy, and assessing service quality based on the classified sentiments. Mahmood *et al*. (2023) extend this line of inquiry by implementing three supervised machine learning classifiers and an ensemble model to determine sentiment polarity (positive, negative, or neutral) in mobile banking reviews. Building on these developments, Akter *et al*. (2024) investigate the role of sentiment analysis in the banking sector, focusing on leveraging customer feedback to enhance service quality and CX. In a comparative analysis of traditional machine learning algorithms and the Bidirectional Encoder Representations from Transformers (BERT) model, they report that BERT significantly outperforms conventional approaches. The model achieves 88% accuracy and an F1 score of 0.86. These results underscore BERT's ability to capture nuanced sentiment patterns. Similarly, Mozumder *et al*. (2024) compare traditional models with BERT for classifying customer feedback into distinct satisfaction drivers in the Fintech industry, again finding that BERT is superior. Additional empirical support for these conclusions is provided by Bhuiyan *et al*. (2024).

Across these studies, a consistent reliance on natural-language inputs, such as reviews, opinions, and other customer-generated texts, is observed. However, analytical outputs remain constrained to predefined sentiment classes, predictive labels, or statistically significant drivers. While these approaches offer benchmarking value, they provide limited insight into how customer experience unfolds during service interactions. Besides, such models struggle to capture value-in-use, relational value, and resource integration, as established by service-dominant logic and relationship marketing theory.

2.2 Conversational agents

Conversational agents represent an attempt to engage customers in interactive service exchanges through which value co-creation and experiential meaning may emerge. In banking, chatbot studies show that customer satisfaction is shaped by interface quality, ease of use, and the chatbot's ability to support the service encounter effectively. This highlights the relevance of these systems for service delivery, while also indicating that much of the evidence remains focused on satisfaction with the interface itself (Eren, 2021). These systems facilitate customer engagement through natural language interactions and, in doing so, function not only as service interfaces but also as valuable sources of real-time data on customer preferences, expectations, and behavioral patterns. However, while these systems can improve customer satisfaction by providing personalized, immediate responses, they may also lead to dissatisfaction if interactions are perceived as impersonal or unhelpful. Trivedi (2019) examines the influence of information quality, system quality, and service quality on CX in banking chatbot interactions, and the subsequent effect on brand love. The study finds that all three quality dimensions significantly shape CX, with system quality encompassing response time, usability, reliability, availability, and adaptability emerging as the most influential factor. Moreover, positive CX is shown to strongly contribute to enhanced brand love. Adam *et al*. (2021) conduct a quantitative analysis of customer interactions with AI-driven chatbots in retail banking, identifying personalization, contextual understanding, and the relevance of responses as critical determinants of chatbot effectiveness and customer satisfaction. Despite their efficiency and scalability, chatbots do not always meet customer expectations.

Qureshi *et al*. (2023) report that customers often perceive chatbots as less effective than human agents or traditional self-service technologies, resulting in lower satisfaction. Petersson *et al*. (2023) analyze chatbot interactions for both simple and complex tasks, finding that human-like features such as personality, politeness, emojis, and helpfulness enhance user experience, while miscommunication errors, especially in simple tasks, negatively affect satisfaction. Le and Nguyen (2024) present survey-based evidence from Vietnam on how chatbot characteristics influence both intrinsic and extrinsic customer values, as well as behavioral outcomes, including satisfaction and continued use. Other regional studies further illustrate chatbot adoption patterns and outcomes. Salem (2024) investigates the role of chatbots in improving CX among 337 banking clients in Gaza, concluding that chatbots enhance CX through continuous availability, rapid responses, and personalized assistance. Park *et al*. (2024) examine factors influencing customer intention to use chatbot services and associated satisfaction levels, finding that anthropomorphism and personalization have significant direct and indirect effects. Makudza *et al*. (2024) employ structural equation modeling on data from 389 Zimbabwean consumers with prior chatbot experience, demonstrating that reliability, responsiveness, and ease of use, especially when integrated into widely used messaging platforms, are key drivers of CX enhancement. Pfoertsch and Sulaj (2023) study chatbot and virtual assistant interactions in Albania and Cyprus, highlighting empathy as a critical factor in AI-driven banking interactions that significantly improves satisfaction and service quality.

Taken together, these studies provide valuable insights into the quality of chatbot–customer interactions, particularly in terms of satisfaction, usability, and relational outcomes. However, the primary emphasis remains on the conversational interface as a service-delivery mechanism and on its immediate outcomes, rather than on systematically capturing the customer experience as it unfolds through repeated interactions with products and services. Consequently, while conversational agents facilitate interaction, they are not designed to interpret, synthesize, and communicate the experiential meanings that underpin long-term relationship development.

2.3 GenAI solutions

Building on the limitations of earlier conversational agents, recent advances in generative AI have opened new possibilities for understanding and enhancing customer experience in banking. Thus, the use of Generative AI in the context of CX in banking remains an emerging but rapidly evolving research domain.

While much of the evidence currently originates from preprints and industry reports, academic contributions are beginning to outline its potential. Fieberg *et al*. (2023), Abdullin *et al*. (2024), Eustaquio-Jiménez *et al*. (2024), and Hettiarachchi (2025) examine the application of models such as GPT-3 and GPT-4 in the banking and service domains. Their studies highlight use cases that include generating synthetic customer dialogs for training service agents, summarizing customer complaints to accelerate issue resolution, and delivering personalized financial advice in real time. In parallel, Landolsi *et al*. (2025) document the adoption of GenAI by banking institutions for generating text, images, voice, and video content to enhance customer engagement.

Despite these technological advancements, research initiatives remain scarce. Few studies specifically employ GenAI to conduct in-depth analyses of customer experience (CX) with the explicit objective of identifying and addressing customer dissatisfaction. This gap highlights a critical direction for future research. It underscores the importance of integrating advanced AI methodologies with conversational and behavioral datasets to generate richer and more actionable insights into customer experiences in the banking industry.

2.4 Objectives of the study

The general objective of this research is multifaceted and is therefore articulated through a set of specific objectives (SO).

2.4.1 SO1: operationalization of service-dominant logic and relationship marketing. Customers engage with financial institutions through multiple digital and non-digital touchpoints while concurrently using banking products and services. These repeated and ongoing interactions collectively constitute the customer experience, which is explicitly enacted across physical branches, digital chats with service agents, and, most saliently, telephone conversations. These interactions represent episodes of value co-creation in which value-in-use emerges through the integration of firm and customer resources. Simultaneously, such recurring interactions contribute to the development of relational value by fostering continuity, trust, and relationship commitment over time.

Building on these premises, this study seeks to surface customer experience as expressed in telephone calls by leveraging competitive learning and generative artificial intelligence (GenAI). By integrating large-scale conversational data with these methods, the proposed analytical framework operationalizes key theoretical constructs, namely value-in-use, resource integration, and relational value, within a real-world banking context. In doing so, the study demonstrates how interaction-level customer experience analytics can support a process-oriented understanding of both value co-creation and long-term relationship development between customers and financial institutions.

2.4.2 SO2: To prioritize contextual interpretation through generative CX extraction. Traditional CX analytics in banking rely on predefined categories. This approach reduces rich conversational data to binary or ordinal outcomes, such as labeling interactions as either “satisfied” or “unsatisfied,” thereby losing important context. To address this limitation, the proposed framework foregrounds contextual interpretation through generative CX extraction. Drawing on advanced natural language processing techniques, it identifies latent themes. It generates coherent, domain-specific representations of customer interactions that reflect not only the expressed content but also the underlying intent, tone, and emotional nuance. By making these interactional meanings accessible, the framework supports a richer understanding of customer experiences as co-created phenomena. It enables non-technical stakeholders to better interpret customer value perceptions without requiring specialized analytical expertise.

2.4.3 SO3: To employ natural language for both input and output. A defining feature of the proposed framework is its exclusive reliance on natural language for both input and output. Consistent with an end-to-end natural language approach, the framework differs from conventional systems that require structured data inputs or produce numerical outputs that require further interpretation. Instead, it directly processes unstructured textual inputs and generates interpretable textual outputs. This design enhances practical usability and accessibility. It enables insights to be readily understood and operationalized by service teams, product designers, and customer experience (CX) strategists, without requiring additional analytical translation.

Additionally, by implementing an entirely natural-language-processing pipeline, the framework addresses a significant gap in current banking practices. Although banks collect large volumes of conversational data through call centers and digital service channels, much of this data remains underexploited due to challenges associated with large-scale text processing and insight extraction. The proposed approach offers an automated, scalable mechanism for converting conversational data into actionable strategic knowledge while preserving the richness and authenticity of customer natural language.

2.4.4 SO4: To develop a novel application of generative AI in banking CX. Generative artificial intelligence (AI) models, such as GPT, have demonstrated strong performance across a range of open-domain tasks, including text summarization, question answering, and content generation. Despite these advances, their application within the financial services sector, particularly for the systematic extraction of customer experience (CX) insights, remains limited. Addressing this gap, the present study develops a novel application of generative AI to analyze and interpret customer interactions in banking contact centers. Specifically, it proposes an intelligent conversational agent designed to operate within the linguistic, procedural, and regulatory constraints characteristic of the banking industry.

The model is trained on domain-specific conversational data, enabling it to generate outputs that maintain semantic fidelity, employ accurate financial terminology, and adhere to institutional communication standards. The resulting CX extractions yield operationally relevant insights, including recurrent service issues (e.g. authentication failures, lost cards, and delayed transfers), emotional friction points (e.g. dissatisfaction with automated services and concerns over unrecognized transactions), and explicit customer requests (e.g. faster resolution times and refund processing). Collectively, these outputs demonstrate the potential of generative AI to support more informed and responsive CX management in banking contexts.

2.4.5 SO5: To create strategies and mitigating measures to make clients self-sufficient. Banks are increasingly pursuing strategies that shift customer interactions from human-agent-assisted channels to digital and self-service alternatives. This shift is intended to improve scalability and reduce the substantial human-resource costs associated with telephone and live chat support. However, persistent customer reliance on direct service agents indicates that existing services and channels often fail to adequately support customers in achieving their goals independently. This reliance not only increases operational costs but also signals unresolved problems, unmet needs, and friction points within current customer journeys. Traditional customer experience analytics provide limited insight into why customers struggle with self-service or repeatedly seek human assistance. The proposed interaction-level generative CX method is therefore needed to capture customer experience in greater depth from contact center calls. In doing so, it reveals the underlying problems, value-in-use, relationship value, needs, and goals that shape customer behavior. By enabling a richer understanding of how customers experience banking services, the method supports the creation of innovations, mitigating measures, strategies, services, new functionalities, and channels that improve existing interactions. These improvements may increase customer satisfaction and loyalty, make customers more self-sufficient, and simultaneously reduce dependence on costly human service resources.

The above objectives mentioned show that the novelty of the proposed framework lies not in any single component taken in isolation, but in its integration into a unified customer-experience analytics pipeline tailored to banking contact-center data. Specifically, the framework combines competitive-learning-based topic organization, the cluster-proportional construction of a retrieval document through an analytical method, and evidence-grounded generative analysis through Hybrid RAG, metadata filtering, semantic re-ranking, and prompt-guided generation. This integration is designed to satisfy the operational requirements for scalability, interpretability, and institutional compliance in regulated financial services environments (An, 2026).

3. Methodology and data

3.1 Method

Banks provide a specific and finite range of products and services, represented by $P = \{p_1, p_{2,\ldots,}p_p\}$. Products are tangible or contract-based offerings that customers can own, access, or enroll in, typically related to money management and may include credit and debit cards, mortgage loans, and investment funds. Services refer to non-physical, value-added activities provided by a financial institution to help customers manage their products, such as making transactions and investing money. This implies that customer needs and complaints concerning products and services are generally well-defined, finite in scope, and prone to recurrence across different customers within a specific context. Moreover, an increase in the number of products and services a customer acquires is typically accompanied by a proportional rise in the corresponding needs and claims (Sedunov, 2020; Birkenmaier and Stratman, 2025).

A customer claim or need expressed in a banking context comprises three core textual components: verbs, functional adjectives, and nouns. Verbs refer to the specific request the customer makes to the institution. It represents the action the customer demands from the bank. For example, "I want to obtain my account statement," "I want to report an unrecognized transaction," or "I would like to open an account." The verbs "obtain," "report," and "open" explicitly express the customer's request. Conversely, functional adjectives, typically past participles such as *lost*, *stolen*, or *blocked*, serve to describe the condition or status of the noun (i.e. products and services) involved in the claim. This helps clarify the customer's problem or situation, giving essential context to their request. Nouns represent the primary objects mentioned in the claim, which, in most cases, correspond to the products and services involved.

Thus, the set of verbs and functional adjectives used by customers in their claims will be defined and finite, represented by $V = \{v_1, v_2, \ldots, v_v\}$ and $C = \{c_1, c_2, \ldots, c_c\}$, respectively.

With the foregoing considerations in mind, the methodology proposed in this study comprises the following stages, shown in Figure 1.

The details of each stage are explained below.

3.1.1 Stage 1-vector representation of conversations. Each phone conversation between agents and customers is represented by a concatenated vector $VCP = concat(V, C, P) \in \{0, 1\}^{v+c+p}$, where V, C, and P encode the presence or absence of verbs, functional adjectives, and nouns, respectively. This representation is invariant to conversation length and word order, as it captures whether a term appears at least once in the interaction.

Let τ denote the set of tokens (words) extracted from a given conversation after preprocessing. Let $\{v_i\}_{i=1}^{v}$, $\{c_i\}_{i=1}^{c}$, and $\{p_i\}_{i=1}^{p}$ denote the vocabularies of verbs, functional adjectives, and nouns terms, defined by the institution respectively. The binary indicator components are defined as follows:

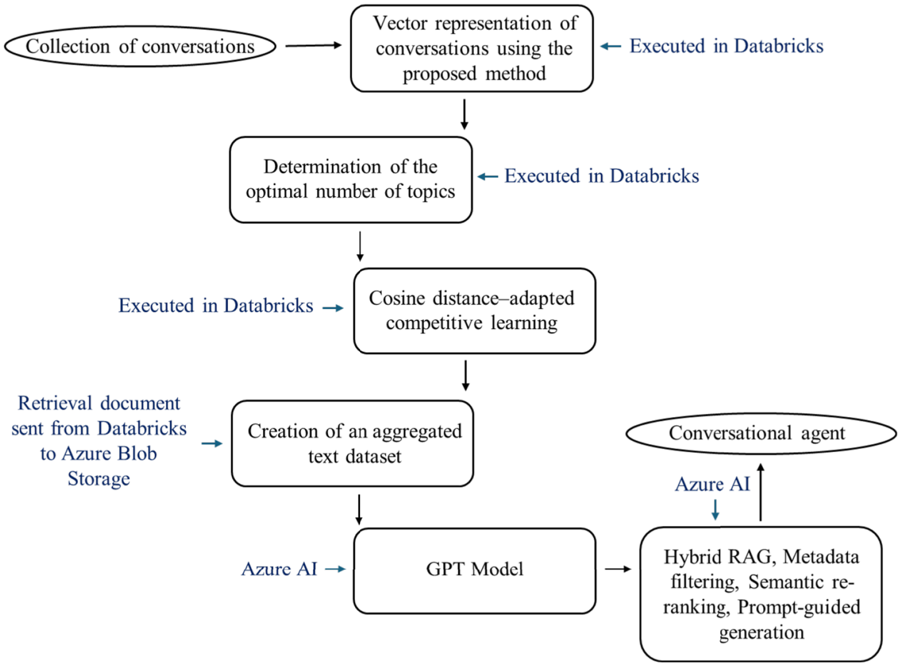

Figure 1. Method flow. Source(s): Authors' own work

$$V_i = f(v_i) = \eta[v_i \in \tau] \, for \, i = 1, 2, \ldots, v \quad (1)$$

$$C_i = g(c_i) = \eta[c_i \in \tau] \, for \, i = 1, 2, \ldots, c \quad (2)$$

$$P_i = h(p_i) = \eta[p_i \in \tau] \, for \, i = 1, 2, \ldots, p \quad (3)$$

Where $\eta[\cdot]$ denotes the indicator function, which equals one if the condition inside the brackets holds and zero otherwise.

Binary vectorization is well-suited for text analysis tasks in which the primary objective is to identify the presence or absence of specific lexical elements while maintaining low computational complexity. Accordingly, the proposed binary lexical vectorization offers a computationally efficient and financially sustainable alternative to transformer-based and embedding-centric methods for large-scale cloud deployment in banking contact center analytics. This approach only encodes whether functionally relevant lexical elements appear in a conversation. Consequently, vector construction scales linearly with the conversation length and can be performed on standard CPU infrastructure, resulting in predictable latency and reduced cloud costs. In contrast, transformer models, such as BERT, rely on self-attention mechanisms with quadratic computational complexity and typically require GPU-enabled resources, which increases the inference time, energy consumption, and financial cost when processing long conversational transcripts (Schwartz *et al*., 2020; Jim *et al*., 2024; Krishnan and Easwarakumar, 2025). The same applies to Top2Vec, which incurs additional overhead due to the generation of dense embeddings and iterative clustering. This makes it more expensive in scenarios involving rolling window updates and frequent retraining (Krishnan and Easwarakumar, 2025). Prior research in service and customer experience analytics underscores the importance of scalable, interpretable, and operationally transparent analytical methods, particularly in regulated financial environments, where cost predictability and continuous execution are essential (Verhoef *et al*., 2021). The proposed representation emphasizes computational efficiency and deployment feasibility, while preserving sufficient discriminative capability for downstream clustering and trend analysis in customer experience research. Recent banking research has likewise shown that text-mining and machine-learning pipelines can uncover service-quality and experience patterns from large-scale customer-generated text, supporting the practical relevance of parsimonious, scalable analytical designs in service analytics (Çallı, 2023).

In this case, the concatenation order *VCP* is used. Still, any order can be applied since this vector is intended to capture the presence or absence of words regardless of their order.

This representation assumes that, for topic discrimination in banking contact-center conversations, the presence of functionally relevant lexical units is more informative than their repetition frequency or precise sequential order. This assumption is appropriate for banking service calls (it should not be indiscriminately generalized to other forms of text, such as books, news articles, or reports), where customers typically contact the center with a relatively specific problem or service need that remains the dominant focus of the interaction. For example, in a call concerning a lost credit card, that issue typically remains the dominant focus of the interaction. However, other topics, such as life insurance, may also arise during the same conversation. In such cases, the vector representation may position the call near others that contain both sets of terms. As a result, the resulting cluster may reflect the primary service issue while also capturing an overlapping service context. For the purpose of this stage, however, this does not materially weaken the analysis, because the objective is to organize conversations by dominant topic for dimensionality reduction and topic structuring, rather than to represent every contextual nuance of the interaction.

Within this setting, term frequency may reflect conversational emphasis, repair, politeness, or call duration rather than a substantively distinct service topic. Likewise, sequential modeling would add substantial representational complexity without being necessary for the clustering objective pursued here. Frequency-based and sequence-aware representations are more suitable when the aim is to model contextual relations, semantic dependencies, and interpretive depth in greater detail. In the proposed framework, these dimensions are intentionally deferred to the GPT stage rather than incorporated into clustering. Conversations are retained in their original natural-language form at the retrieval-document stage, allowing the subsequent generative model to examine contextual meaning, inter-word relationships, and problem interpretation more comprehensively.

3.1.2 Stage 2- cosine distance–adapted competitive learning. Once each conversation is represented as a concatenated binary vector, a strict competitive learning algorithm is implemented, incorporating specific adaptations to utilize the cosine distance. This metric is particularly appropriate for assessing the similarity between two non-zero vectors because it quantifies their angular separation while remaining invariant to differences in magnitude. In addition, in clustering normalized binary textual representations, where each vector encodes the presence or absence of words, cosine distance is preferred over Euclidean distance. This is because of its theoretical alignment with normalized overlap, resistance to text-length bias, and superior discriminative power in sparse high-dimensional spaces. Empirical studies, including binary similarity surveys (Muniswamaiah *et al.*, 2023; Garcia-Morato *et al.*, 2025) and silhouette-based clustering evaluations (Moufok *et al.*, 2025), demonstrate that cosine distance outperforms Euclidean metrics, particularly in variable-length, sparse input contexts. Furthermore, recent analytical work by Ismael (2025) emphasizes the limitations of Euclidean distance in such data regimes and underscores cosine interpretability and scale-invariance. Thus, for topic-based text clustering with binary presence vectors, cosine distance is both theoretically grounded and empirically validated.

The objective of competitive learning is to cluster conversations by common topics (a topic may encompass multiple issues across several products and services) for dimensionality reduction, that is, to represent a set of m conversations using K topics. This approach is particularly suitable in the banking sector, where groups of customers often discuss recurring issues such as “lost credit card,” “blocked account,” or “malfunctioning mobile application.”

Competitive learning postulates that Kneurons engage in mutual competition to become activated by a given input (Du, 2010). In the present study, the winning neuron is selected using a cosine-distance criterion. Specifically, the neuron exhibiting the lowest cosine distance to the input is designated as the winner and thereby acquires that input as part of its cluster. Consequently, at each iteration i, the neurons seek to minimize the cost function defined in Equation (4) to secure the input x_i.

$$E_i = \sum_{k=1}^{K} \mu_{k,i} \left(1 - \frac{x_i^T c_k}{\|x_i\| \, \|c_k\|} \right) \tag{4}$$

Where $\mu_{k,i}$ is equal to 1 if the input $x_i \in \mathbb{R}^d$ is won by neuron $c_k \in \mathbb{R}^d$ and 0 otherwise. The expression $1 - \frac{x_i^T c_k}{\|x_i\| \, \|c_k\|}$ is the cosine distance, that in general terms, ranges from 0 (most similar) to 2 (most dissimilar). To become more competitive, neuron c_k must move in the direction where the cost function E_i decreases most rapidly. This direction is determined by the negative gradient of the function E_i with respect to c_k, as shown in Equation (5).

$$-\frac{\partial E_i}{\partial c_k} = -\sum_{k=1}^{K} \mu_{k,i} \frac{\partial}{\partial c_k} \left(1 - \frac{x_i^T c_k}{\|x_i\| \, \|c_k\|} \right) \tag{5}$$

If the vectors x_i and c_k are normalized to unit length, that is, $\|x_i\| = 1$ and $\|c_k\| = 1$, Equation (5) is simplified, as shown in Equations (6) and (7).

$$-\frac{\partial E_i}{\partial c_k} = -\sum_{k=1}^{K} \mu_{k,i} \frac{\partial}{\partial c_k}\left(1 - x_i^T c_k\right) \quad (6)$$

$$-\frac{\partial E_i}{\partial c_k} = -\sum_{k=1}^{K} \mu_{k,i} \left(-x_i\right) \quad (7)$$

If c_w (with $w \in \{1, 2, ...K\}$) is the winner, then $\mu_{w,i} = 1$ and $\mu_{k,i} = 0$ for the rest of the neurons where, $k \neq w$, and the gradient is simplified, as shown in Equation (8).

$$-\frac{\partial E_i}{\partial c_w} = x_i \quad (8)$$

Equation (8) indicates that, at each iteration, the neurons adjust their positions towards the inputs to minimize the cost function and increase their likelihood of winning. Substituting the gradient into the conventional learning rule employed in both supervised and unsupervised machine learning models (Tian *et al.*, 2023) yields the following expression:

$$c_w(i+1) = c_w(i) - \alpha \frac{\partial E_i}{\partial c_w} = c_w(i) - \alpha(-x_i) \quad (9)$$

Where α denotes the scalar learning rate, which is typically a small positive value. Thus, the learning rules for the K competing neurons are defined by Equations (10) and (11).

$$c_w(i+1) = c_w(i) + \alpha x_i \quad (10)$$

$$c_k(i+1) = c_k(i) \, for \, k \neq w \quad (11)$$

The gradient of the cosine distance is determined by angular relationships rather than vector magnitudes. While generally stable, such gradients may exhibit sparsity or variability across different dimensions. The Adaptive Moment Estimation (Adam) algorithm is particularly well-suited to these conditions, as it simultaneously estimates the first moment (the gradient mean) and the second moment (the gradient variance). This dual estimation enables balanced and adaptive parameter updates (Défossez *et al.*, 2022). It is particularly effective in high-dimensional spaces where certain features may exert a disproportionate influence on the optimization process. Consequently, integrating Adam into the learning rules specified in Equations (10) and (11) produces the outcomes reported in Equations (12) and (13).

$$c_{w(i+1)} = c_w(i) - \alpha \frac{\widehat{m}_i}{\sqrt{\widehat{v}_i} + \in} \quad (12)$$

$$c_k(i+1) = c_k(i) \, for \, k \neq w \quad (13)$$

Where $\widehat{m}_i = \frac{\beta_1 m_{i-1} + (1-\beta_1) g_i}{1-\beta_1^i}$ and $\widehat{v}_i = \frac{\beta_2 v_{i-1} + (1-\beta_2) g_i{}^2}{1-\beta_2^i}$ are the exponential moving average of gradients and squared gradients, respectively with bias correction. With $g_i = \frac{\partial E_i}{\partial c_k}$, α as the learning rate, ϵ is a small constant to avoid division by zero, and $\beta_1, \beta_2 \in [0, 1)$. After each update, the winning prototype is renormalized to unit length to preserve the cosine-distance geometry.

The objective of these learning rules is to identify the optimal centroids that cluster the conversation vectors into k principal topics.

Compared with embedding-based topic modeling approaches such as BERTopic and Top2Vec, the proposed competitive-learning method offers clear advantages in scalability, transparency, and institutional compliance, which are critical in banking environments. BERTopic and Top2Vec rely on external public libraries and pre-trained embedding models that function as black-box components and are subject to versioning, licensing, and security constraints. In contrast, the proposed approach is implemented using native mathematical operations, without dependence on third-party NLP frameworks, and therefore aligns with strict bank security and data-governance policies. The learning dynamics are explicitly defined through analytical cost functions and gradient-based update rules, providing complete transparency of the optimization process. Unlike BERTopic and Top2Vec, where adaptive optimization is implicit in pre-trained models and not accessible at the topic-learning stage, the proposed method explicitly incorporates adaptive learning rules, such as Adam. This explicit formulation enables controllable, auditable, and deterministic topic formation. It also supports linear scalability to large volumes of customer-service conversations and efficient daily execution under constrained computational budgets. By avoiding repeated transformer inference and opaque optimization layers, the method reduces computational and energy consumption. The latter supports banks' sustainability objectives and the long-term maintainability of customer experience analytics pipelines (Shreyashree *et al.*, 2022; Mohd Amin *et al.*, 2025; Yeo *et al.*, 2025).

3.1.3 Stage 3- determination of the optimal number of topics and coherence. The algorithm is designed to determine the optimal value of K from the set of m representative input vectors, as shown in Figure 2. The latter is accomplished by treating the number of clusters as a parameter to be estimated, selecting the value that maximizes the Calinski–Harabasz (CH) index. The CH index is a widely adopted metric for evaluating clustering quality. It measures the ratio of between-cluster dispersion to within-cluster dispersion, thereby favoring solutions characterized by high intra-cluster compactness and clear inter-cluster separation (Qiu *et al.*, 2020; Farea *et al.*, 2024). From a computational perspective, the CH index is considered relatively efficient and low in complexity, particularly compared with alternative clustering validation metrics such as the Silhouette coefficient.

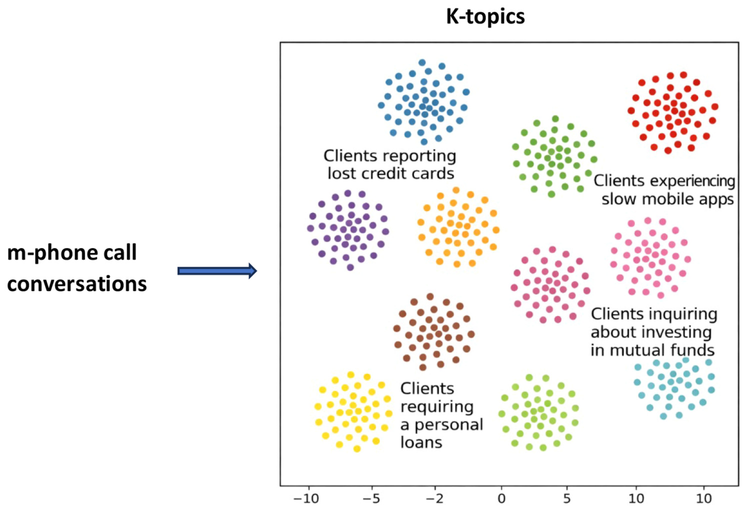

Figure 2. K-topics from m-phone call conversations. Source(s): Authors' own work

The proposed method is expected to yield coherent topics from banking contact center data because it is explicitly aligned with the linguistic and operational structures of service interactions. Banking conversations are not free-form text; instead, they are lexically constrained and dominated by operational verbs (e.g. *block, transfer, reverse, cancel, activate*), functional adjectives (e.g. *pending, fraudulent, overdue, authorized*), and domain-specific nouns (e.g. *card, charge, account, balance, limit*). By restricting textual representations to these parts of speech, the method induces lexically tight topic boundaries. This is further confirmed by computing the C_v metric, which ensures that the generated topics and their associated conversations exhibit actual coherence. C_v topic coherence evaluates topic interpretability by quantifying the extent to which a topic's most representative terms co-occur within a reference corpus. The metric integrates normalized pointwise mutual information (NPMI), sliding-window co-occurrence statistics, and cosine-similarity aggregation over word context vectors. Higher C_v values indicate semantically coherent topics whose term groupings align closely with human judgments. As a result, C_v is widely used to compare topic models and to guide the selection of the number of topics in applied text-mining research (Mahdikhani and Meena, 2024).

Although individual conversations may involve multiple purposes or intents, the competitive-learning stage is designed to capture the dominant interactional context underlying contact-center demand for each topic. As a result, a given topic can contain different service requests and resolution patterns that co-occur within the same conversation. After topic assignment, the retrieval stage does not transform or decompose the conversations into separate intent units; instead, the original transcripts are retained in natural language and used directly for plain-text retrieval. Consequently, when a conversation contains multiple intents, these are represented jointly within the cluster to which the conversation is assigned, rather than being split across multiple topics. This hard-assignment strategy may introduce some degree of topic dilution in especially complex cases, but it remains consistent with the study's objective of identifying the main drivers of demand at scale and dimension reduction. At the same time, secondary issues are not lost, because they remain embedded in the original transcripts and can still be accessed during the downstream retrieval and generative analysis stages, even when the conversation is mapped to a single winning neuron.

3.1.4 Stage 4- creation of an aggregated text, the retrieval document. Once the conversations have been clustered according to their representative vectors, a proportion r of random samples is drawn from each cluster. These samples are then combined to form the retrieval plain-text dataset used by the RAG process. This approach accounts for the frequency effect of the topics addressed in the contact center. For instance, if a larger number of customers contact the center to report a lost credit card, whereas only a few inquire about insurance purchases, more conversations will be extracted from the former topic. This ensures that it carries greater weight in the final aggregated text, which compiles all sampled conversations.

3.1.5 Stage 5- GPT model with hybrid RAG, semantic re-ranking, metadata filtering, and prompt. The GPT model processes the retrieval document generated in the previous step by incorporating these four components.

(1) Hybrid Retrieval-Augmented Generation (Hybrid RAG). Hybrid RAG defines how candidate conversational evidence is retrieved from the plain-text corpus. It combines BM25 lexical retrieval with semantic vector similarity in Azure AI Search. This captures both exact banking terminology and semantically similar customer expressions. The approach maximizes recall and grounding at the retrieval stage.

(2) Metadata filtering. Metadata filtering is applied at retrieval time using structured fields in Azure AI Search. It restricts which subsets of the plain-text corpus are eligible for retrieval using structured attributes such as product category, channel, jurisdiction, period, and topic. It does not analyze text content; instead, it ensures that only

contextually comparable and institutionally valid conversations are used as evidence. In CX analysis, this grounds insights into the correct service context and regulatory scope.

(3) Semantic re-ranking. After hybrid retrieval, candidate conversations are re-ranked using Azure AI Search's semantic ranking capabilities. It re-orders candidate conversations based on contextual relevance to the user's analytical query. This step improves precision by prioritizing conversations that most directly reflect the customer experience dimension of interest (e.g. dissatisfaction, effort, resolution quality, pain points, emotions), while demoting marginal or tangential interactions.

(4) Prompt-guided generation. Prompt-guided generation governs how retrieved conversational evidence is synthesized. A structured, evidence-constrained prompt instructs the GPT model to extract predefined customer-experience insights from the retrieved text. It answers predefined questions on satisfaction, sentiment, service quality, loyalty, and product interactions. This produces interpretable, decision-relevant insights rather than free-form summaries.

Hybrid RAG, metadata filtering, semantic re-ranking, and prompt-guided generation are four interdependent components that operate sequentially on the plain-text retrieval corpus. Together, they ensure that customer insights are representative, contextually valid, semantically precise, and grounded in banking contact center conversations. At the same time, computationally intensive algorithms are avoided, thereby lowering operational costs.

Such a design is also aligned with recent evidence that AI-enabled service systems in banking must remain controllable and governance-aware. This is particularly important because operational deployment unfolds amid socio-technical and regulatory tensions that shape how AI can be integrated into frontline service processes (Sathiskumar and Andersen, 2025).

3.2 Data

The dataset analyzed in this study derives from audio recordings of interactions between customers and service agents at the contact center of a major financial institution with a global presence. To demonstrate the functioning and accuracy of the proposed methodology, only conversations from one English-speaking country where the institution operates are considered. These recordings capture authentic banking service scenarios, including account-related inquiries, transaction dispute resolution, product consultations, and complaint handling. The dataset comprises monthly recordings collected between July 2023 and July 2025, totaling 25 months, and encompassing 10,062,406 calls. For confidentiality and in accordance with the institution's privacy policies, the bank's name cannot be disclosed.

3.2.1 Data extraction. The dataset is derived from call recordings captured through the institution's contact center telephony system. Each interaction is stored in a standardized digital audio format, typically as compressed files (e.g. MP3 or OGG). Structured metadata, including call identifier, timestamp, duration, and agent identifier, also accompanies it. The extraction procedure involves securely transferring the audio files from the call management repository to a controlled private cloud environment. All transfers adhere strictly to the institution's internal security protocols and to relevant financial and data protection regulations. This ensures data integrity and confidentiality.

3.2.2 Audio-to-text conversion. Following extraction, the audio files were converted to text using speech-to-text processing, producing transcripts suitable for analysis. This process is implemented through a hybrid processing pipeline combining Python and PySpark, enabling efficient large-scale transcription. The conversion workflow comprises the following sequential stages:

Audio Standardization – All recordings are converted to a consistent format (WAV, 16 kHz, mono channel) to ensure compatibility and optimize transcription accuracy. Standardization is performed using the Python libraries *pydub* and *librosa*.

Segmentation – Segments of audio containing promotional messages, automated announcements, and survey prompts are removed to enhance recognition accuracy and reduce memory consumption. The segmentation process employs *pydub* for local processing and *pyspark.ml.feature* utilities for distributed large-scale execution.

Noise Reduction – To mitigate background noise common in contact center environments, spectral noise suppression techniques are applied using the *librosa* and *noisereduce* libraries in Python.

Speech Recognition – Transcription is conducted using a multi-engine strategy. For small-scale datasets, the *speech_recognition* library in Python is employed to enable efficient synchronous processing. For large-scale datasets, the OpenAI Whisper API and the Vosk speech recognition engine are integrated into PySpark workflows, enabling distributed, parallel execution across multiple nodes. This architecture ensures high transcription accuracy while maintaining the scalability needed to process large volumes of conversational data.

3.2.3 Masking of sensitive information. To comply with applicable privacy legislation and institutional confidentiality requirements, all personally identifiable information (PII) is systematically anonymized before analysis. An anonymization pipeline is implemented using a hybrid Named Entity Recognition (NER) approach that combines rule-based methods with machine learning techniques. NER, a core methodology in NLP, enables the detection and classification of entities within unstructured text, assigning them to predefined categories such as personal names, organizations, dates, and numeric identifiers.

Detected sensitive elements are substituted with standardized tokens as follows:

PHONE – telephone numbers (mobile or landline)

ADDRESS – physical or mailing addresses

ACCOUNT – bank account numbers

EMAIL – email addresses

ID_NUMBER – government-issued identification numbers

CARD_NUMBER – debit or credit card numbers

DOB – dates of birth

AMOUNT – monetary amounts linked to specific accounts

COMPANY – any company or organization name mentioned

NAME – any person's name mentioned in the conversation

The masking protocol combines regular expressions for pattern-based detection with domain-specific dictionaries to capture banking-related terminology. It also leverages pre-trained NER models from spaCy and Hugging Face's transformer-based architecture, fine-tuned on financial domain corpora. This integrated approach achieved high recall, capturing the vast majority of PII instances, while maintaining high precision to minimize false positives.

All libraries and modules referenced in this section for data processing are implemented in the cloud environment using Python or PySpark.

3.2.4 Text normalization. Following transcription and anonymization, all text is normalized to ensure consistency and suitability for downstream NLP tasks. The normalization pipeline involves:

Case Standardization – Converting all text to lowercase.

Token Cleaning – Removing extraneous punctuation, transcription artifacts, and non-linguistic symbols.

Stopword Removal – Removing non-informative words through an internally defined set of stopword lists established according to expert criteria. For example, when the customer pauses to think and produces expressions such as "uh" or "mmh," or when laughter occurs, which is represented in the transcript as "haha."

The final dataset consists of fully anonymized and normalized textual transcripts. Each transcript is associated with metadata that includes anonymized agent and customer identifiers, the call date, and the call duration. No original audio files or unmasked transcripts are retained. Table 1 presents an example of how the original conversation transcripts are processed to remove as much noise as possible. The processed transcript is then used as input for the proposed methodology.

The anonymization procedure shown in Table 1 protects sensitive data while preserving the nature and context of the original information, which is essential to the proposed methodology. For the methods and the extraction of customer experience insights, specific details such as customer addresses with street and number are not required. Similarly, it is unnecessary to retain the exact name of an online store where a product was purchased or the personal names of customers or agents. The model requires only general indicators, such as a word or key phrase, to identify whether an address, an online store, or a proper name is mentioned.

A limitation of this pipeline is that speech-to-text errors may propagate into lexical vectorization, particularly when operational verbs or functional adjectives are misrecognized. This risk was mitigated through audio standardization, segmentation, noise reduction, multi-engine transcription, and subsequent text normalization. Moreover, because topic assignments are driven by recurring lexical patterns across large volumes of conversation rather than by isolated token occurrences, the effect of occasional transcription errors is attenuated, though not fully eliminated.

Table 1. Transcript text processing

Original transcript	Processed transcript
Agent: Good morning, thank you for calling Bank Name customer service, my name is Laura, how can I help you today?	*agent:* good morning thank you for calling [company] customer service my name is [name] how can i help you today
Client: Hi, uh, yeah, I think I lost my credit card yesterday and I don't know what to do	*client:* hi yeah i think i lost my credit card yesterday and i dont know what to do
Agent: I'm sorry to hear that. Don't worry, we'll take care of it. Can I have your full name and the last four digits of your account, please?	*agent:* im sorry to hear that dont worry well take care of it can i have your full name and the last four digits of your account please
Client: Sure, it's Carlos Hernández, and uh, the last four are 4,923	*client:* sure its [name] and the last four are [account]
Agent: Thank you, Mr. Hernández. Before we continue I'll need to verify your identity. Could you confirm your date of birth and the billing address on file?	*agent:* thank you [name] before we continue ill need to verify your identity could you confirm your date of birth and the billing address on file
Client: Yeah, it's October 14, 1986, and the address is 125 Avenida Central, uh . . . Monterrey	*client:* yeah its [dob] and the address is [address]
Agent: Perfect, thank you. I've verified your account. So you mentioned the card was lost yesterday, right?	*agent:* perfect thank you ive verified your account so you mentioned the card was lost yesterday right
Client: Yes, I think maybe I left it at the restaurant or taxi, I'm not really sure	*client:* yes i think maybe i left it at the restaurant or taxi im not really sure
Agent: Understood. For your security, I'll go ahead and block this card immediately so no one else can use it	*agent:* understood for your security ill go ahead and block this card immediately so no one else can use it

Source(s): Authors' own work

4. Results and discussion

4.1 General results

A conversational agent trained using the proposed method is made available to the institution's customer experience analysts, enabling them to extract insights through natural language queries. The backend of this conversational agent continuously receives telephone conversations between customers and contact center agents.

Topic extraction is implemented through a competitive learning process operating on a rolling one-month window of historical data, updated daily. This historical window spans 31 calendar days and represents the minimum temporal horizon required to capture key operational and contextual dynamics in banking contact centers. These dynamics include billing cycles and monthly fee assessments (biweekly or end-of-month), card replacements and fraud-related incidents (typically resolved within 7–15 days in the institution examined), regulatory or policy changes (generally enforced at the beginning of a calendar month), application updates or service outages (commonly deployed on a monthly release cycle), and marketing campaigns and promotional activities (refreshed monthly). Collectively, these factors have been shown to shape the volume and nature of customer interactions with the contact center (Voorhees *et al*., 2017; Zhang *et al*., 2023; Shamsuzzoha *et al*., 2025). Although a longer temporal window could also incorporate these operational and contextual effects, such an extension is not strictly necessary for effective topic extraction. It would introduce additional computational overhead and associated costs without a commensurate gain in representational relevance. Hourly or sub-daily updates are not adopted because the semantic content of customer interactions in bank contact centers evolves over multi-day horizons rather than within hours. Intraday variability mainly reflects fluctuations in interaction volume rather than substantive changes in customer issues. Prior research in service operations and analytics shows that meaningful shifts in customer concerns are driven by operational events, service disruptions, and policy interventions whose effects persist over several days. Persistent model updates tend to increase noise, reduce topic stability, and introduce unnecessary computational overhead without improving interpretability (Pacella *et al*., 2024; Ranieri *et al*., 2024; Rahman *et al*., 2022).

After the topics and their corresponding conversations are established, a retrieval plain-text corpus is constructed by randomly sampling a proportion r of conversations from each topic. This corpus is subsequently used as input for the GPT model, integrating hybrid RAG, semantic re-ranking, metadata filtering, and prompt-guided generation in Azure AI.

Thus, the optimal number of topics and length of the plaintext corpus may vary daily, depending on the nature and context of incoming conversations. This approach ensures that queries submitted to the chatbot yield customer experience (CX) insights that reflect the most current information.

As mentioned previously, the model is executed daily in practice. To demonstrate the performance of the proposed method, this study reports results from the last day of each month from July 2023 to July 2025. The analysis uses all conversations collected up to that date, totaling 25 months and 10,062,406 conversations.

Figure 3 presents the total of conversations analyzed, classified into service requests and problem resolution. Overall, service-related calls are more frequent than problem resolution calls. A slight upward trend is observed, attributed to the natural increase in the number of clients and contracted products. Importantly, this trend is undesirable, as it implies greater demand on contact centers, a greater need for service agents, and longer customer waiting times. Ultimately, this hurts both customers and institutions.

Figure 4 illustrates the distribution of conversation frequencies based on word count. For the institution under analysis, most conversations range from approximately 500 to 1,000 words, with a progressively lower incidence of more prolonged interactions. In some cases, customer issues are not resolved by the initial agent and are escalated to subsequent agents, leading to more extended conversations. These prolonged interactions are also retained in the analysis.

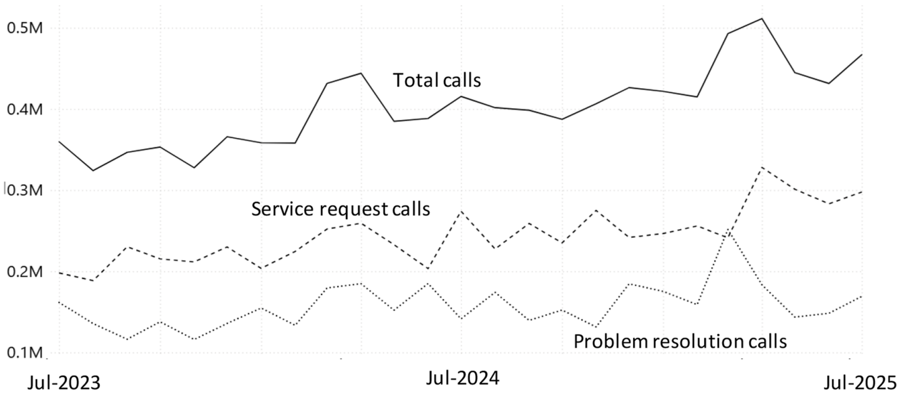

Figure 3. Number of conversations. Source(s): Authors' own work

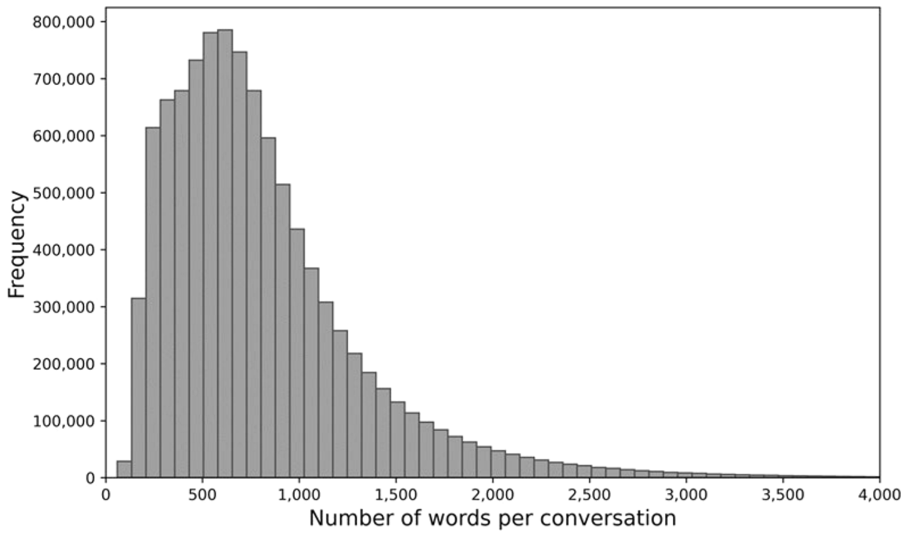

Figure 4. Histogram of number of words per conversation. Note: Bank contact center data from July 2023 to July 2025. Source(s): Authors' own work

The competitive learning algorithm is provided with a range of 1–200 neurons, allowing it to autonomously determine the optimal number of topics based on the Calinski–Harabasz (CH) index. Figure 5 presents the optimal number of topics across the entire study period, which shows a modest upward trend. This trend reflects the institution's introduction of new products and services, which broadens the range of issues that lead customers to contact the service center. In the absence of mitigating measures, this growth reflects the business's natural expansion.

It is observed that, over the analysis period, the optimal number of topics fluctuates between 89 and 98. This relatively narrow fluctuation band likely reflects a combination of substantive changes in customer demand and normal sensitivity of the Calinski–Harabasz

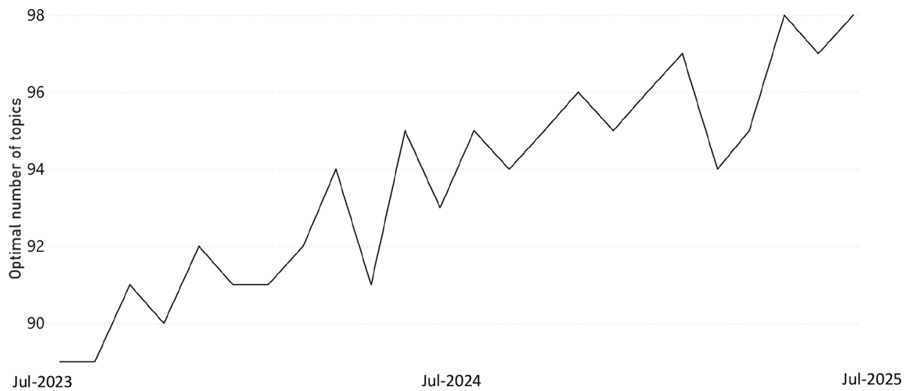

Figure 5. Optimal number of topics for July 2023–July 2025. Source(s): Authors' own work

index under rolling-window updates. The absence of abrupt swings suggests that the institution's topic structure remains broadly stable over time, while modest month-to-month variation captures both incremental business change and expected index responsiveness to updated conversational mixes.

Table 2 presents a sample of the topic classifications identified in July 2023. Of these, 55 were related to customer service and 34 to problem resolution, corresponding to approximately 61% and 39%, respectively. The topic titles are determined through a prompt sent to the GPT model via an API in Databricks once the clusters have been created. It is important to note that although the results for July 2023 are presented here, the topic titles

Table 2. Main topics for July 2023

Service request					
Account balance check	Request for account statements	Request for international money transfer	Enrollment in loyalty or rewards program	Request for premium or exclusive banking services	Request for financial advisory or consultation
Inquiry on recent transactions	Addition of new beneficiaries	Setup of wire transfer instructions	Redemption of reward points	Request for promotional offers or discounts	Request for demonstration of internet banking services
Problem resolution					
Unsuccessful ATM cash withdrawal	ATM debited account without dispensing cash	Declined payment at point of sale or online	Dispute regarding unauthorized transaction	Report of lost debit card	Report of lost credit card
Report of stolen card	Mobile banking application malfunction	Inability to log into online banking	Account access locked	Display of incorrect account balance	Duplicate debit transaction

Source(s): Authors' own work

remain broadly stable over time due to the bank's contact center's limited lexicon and constrained context.

4.2 Benchmarking results with other methods

Table 3 reports the results for the C_v metric, execution time, and execution cost of the proposed methodology, along with comparisons with the BERTopic and Top2Vec methods. Only five periods were selected for executing BERTopic and Top2Vec and for comparing the aforementioned metrics, due to the budgetary constraints of this study.

BERTopic and Top2Vec are prominent embedding-based approaches to topic modeling that rely on dense semantic representations, typically derived from transformer-based embeddings or joint document–word vector spaces, to uncover latent thematic structures in large-scale unstructured text corpora. These methods are widely adopted in exploratory semantic analyses across domains such as social media analytics, customer feedback mining, and marketing research, where semantic expressiveness and modeling flexibility are often prioritized over strict probabilistic assumptions.

Despite their widespread adoption, prior research highlights that these embedding-based text analytics methods are primarily employed for exploratory insight generation, while posing challenges related to transparency, reproducibility, cost, execution time, and governance when deployed in operational or regulated environments. Consequently, methodological choices remain highly context-dependent, and organizations frequently complement or benchmark such approaches against more controllable and auditable analytical frameworks (Steidl *et al*., 2023; Chen *et al*., 2026).

The C_v metric remains above 0.6 for the proposed methodology across all months of the analysis period. This level is considered acceptable by the institution for topic coherence. It also slightly exceeds the C_v values obtained by BERTopic and Top2Vec, which remain below 0.6 across the five selected executions. The lower C_v coherence values observed for BERTopic and Top2Vec may be due to some dilution in the embedding space, with topics showing more variance and being sensitive to heterogeneous conversation length and content. These models group lexically diverse but semantically related terms, thereby weakening local word co-occurrence patterns and reducing co-occurrence-based coherence. By contrast, the proposed method restricts representations to action-oriented and domain-specific lexical units. It also enforces competitive clustering. As a result, topics are lexically compact, and their representative terms consistently co-occur in customer–agent interactions.

In the present study, C_v coherence is interpreted primarily as an indicator of operational interpretability: higher values suggest that the constrained lexical units defining each topic form recognizable and managerially meaningful issue groupings within the banking contact-center context. It should therefore be understood as evidence of an interpretable topic structure under the proposed representation, rather than as an exhaustive measure of the full semantic relationships present in the conversations.

Regarding execution time, a significant difference is observed between the proposed method and the other two approaches across the five executions under the same runtime settings. The proposed methodology can be updated without difficulty every 24 h. In contrast, the other methods require more than 24 h to produce updated results. These execution times directly affect execution costs, which are clearly more efficient under the proposed methodology. BERTopic and Top2Vec are slower at scale because they rely on dense embedding inference and high-dimensional similarity operations, whose costs grow with both corpus size and conversation length. The proposed method avoids embedding generation and instead uses sparse lexical representations with analytically defined competitive learning, resulting in lower computational overhead and faster convergence.

Because BERTopic and Top2Vec were run over a limited number of periods, the benchmarking results should be interpreted with caution. In this study, these comparisons are intended primarily to assess computational efficiency, execution cost, and operational

Table 3. Comparison of metrics across different methods

Month	Number of conversations	Proposed method (C_v metric)	BERTopic (C_v metric)	Top2Vec (C_v metric)	Proposed method (hours)	Proposed method cost (USD)	BERTopic (hours)	BERTopic cost (USD)	Top2Vec (hours)	Top2Vec cost (USD)
2023–07	3,59,658	0.610			6	60				
2023–08	3,24,064	0.609			6	57				
2023–09	3,46,632	0.612			6	58				
2023–10	3,53,193	0.600			7	61				
2023–11	3,27,631	0.608			6	54				
2023–12	3,65,828	0.611			7	62				
2024–01	3,58,369	0.614			7	62				
2024–02	3,58,070	0.613			7	62				
2024–03	4,31,662	0.617			7	71				
2024–04	4,44,053	0.615			8	72				
2024–05	3,84,893	0.618			6	69				
2024–06	3,88,355	0.616			7	70				
2024–07	*4,15,542*	*0.615*	*0.571*	*0.533*	*7*	*71*	*46*	*451*	*64*	*638*
2024–08	4,01,638	0.613			7	61				
2024–09	3,98,512	0.611			7	63				
2024–10	*3,87,253*	*0.612*	*0.575*	*0.533*	*6*	*59*	*41*	*410*	*58*	*615*
2024–11	4,06,447	0.619			7	71				
2024–12	4,26,391	0.614			7	70				
2025–01	*4,21,831*	*0.615*	*0.573*	*0.532*	*7*	*69*	*47*	*463*	*65*	*668*
2025–02	4,14,878	0.614			7	71				
2025–03	4,92,851	0.616			8	84				
2025–04	*5,11,469*	*0.617*	*0.574*	*0.537*	*9*	*85*	*58*	*597*	*78*	*791*
2025–05	4,44,734	0.615			8	72				
2025–06	4,31,490	0.613			7	76				
2025–07	*4,66,962*	*0.618*	*0.576*	*0.536*	*8*	*72*	*52*	*516*	*71*	*703*

Note(s): Metrics for the last day of the month. Execution time values and cost may vary from one run to another and across institutions, as cloud-based tools are continuously updated. Execution times and costs are rounded to the nearest integer

Source(s): Authors' own work

feasibility in a regulated banking environment, rather than to claim definitive superiority across all topic-modeling contexts.

4.3 Validation

To evaluate the accuracy of the customer experience insights extraction, fifteen key business questions are formulated, each corresponding to one of three strategic business areas: satisfaction and emotional sentiment, perceived quality and experience, and preferences and loyalty indicators.

Satisfaction and emotional sentiment

(1) Do customers express satisfaction after their issue is resolved by the contact center?

(2) What emotions (e.g. frustration, relief, gratitude) are most frequently expressed during interactions?

(3) Are there specific products or services that consistently generate positive emotional responses?

(4) Which types of issues are most commonly associated with customer frustration or dissatisfaction?

(5) Do customers express a sense of trust and confidence in the bank after speaking with an agent?

Perceived quality and experience

(6) How do customers describe the overall quality of service received during the call?

(7) What aspects of the interaction (e.g. agent attitude, clarity, speed) contribute most to customer satisfaction?

(8) Do customers feel that their concerns were listened to and understood by the agent?

(9) Which moments during the call are most likely to trigger dissatisfaction or complaints?

(10) Is there evidence of customers feeling valued or appreciated during the conversation?

Preferences and loyalty indicators

(11) Do customers express preferences for specific channels (phone, app, in-branch) based on previous experiences?

(12) What factors make customers feel loyal to the bank or likely to recommend it to others?

(13) Are there indications that a customer is considering switching to a competitor due to dissatisfaction?

(14) What product features or service attributes are most frequently associated with customer happiness?

(15) Do customers express willingness to use additional products or services based on their current satisfaction?

The defined questions are submitted to the developed model, and the responses are evaluated using two complementary validation approaches. The first relies on human intelligence for rubric validation, while the second involves independent validation using alternative AI methodologies.

4.3.1 Rubric validation by human intelligence. The inclusion of human evaluators is theoretically and practically justified. Generative AI systems often exhibit hallucinations and overconfidence in unsupported claims, which cannot be reliably detected by automated metrics. Moreover, in customer experience contexts, qualitative dimensions such as tone, empathy, and compliance with privacy norms are indispensable yet inherently subjective. As emphasized in prior studies, human judgment remains the gold standard for evaluating contextual quality, particularly in domain-specific applications (Fan *et al*., 2025; Hashemi *et al*., 2024).

This approach also aligns with emerging calls in generative AI research to adopt rubric-based evaluations that explicitly encode human values and domain constraints (Hashemi *et al*., 2024).

The rubric validation proposed defines accuracy as a multidimensional construct operationalized through five criteria: Correctness, Faithfulness, Completeness, Conciseness, and Compliance (Novotna and Libal, 2022). Each response is rated on a 0–5 ordinal scale, where 0 denotes non-compliance, and 5 reflects full compliance. This approach is consistent with established best practices in natural language generation (NLG) evaluation. These practices recommend decomposing quality into interpretable dimensions rather than relying solely on single scalar metrics (Fan *et al*., 2025). The final score for each GPT-generated answer is computed as the average across the criteria, enabling a balanced assessment of factual, structural, and ethical performance.

The description of each criterion is as follows:

Correctness ensures factual reliability, consistent with factual grounding frameworks (Hashemi *et al*., 2024).

Faithfulness guards against hallucinations by requiring direct support from the underlying corpus of customer–agent conversations.

Completeness captures responsiveness to the entire query, addressing concerns of partial or selective answering.

Conciseness mitigates verbosity and ensures information is delivered efficiently.

Compliance enforces regulatory and ethical safeguards, particularly against disclosure of personally identifiable information (PII), aligning with financial-sector data protection norms.

Human validation is designed as a distributed expert evaluation framework rather than a co-rating scheme. Thirty trained contact center agents with domain expertise in interpreting customer–agent interactions conduct the evaluations each month. At the beginning of every month, the agents are assigned to validate GPT-generated responses to 15 predefined analytical questions. Each agent reviews a randomly sampled subset of 50 customer–agent conversations drawn without replacement from the full monthly corpus, resulting in an evidence base of 1,500 conversations per month. Each set of 50 conversations is used to construct the plain-text retrieval that feeds the GPT model to generate the responses based on that sample in the demo section of Azure AI. This process is repeated over 25 consecutive months (July 2023–July 2025), yielding a total of 37,500 distinct conversations (30 agents $\times$ 50 conversations $\times$ 25 months) reviewed as contextual evidence and 11,250 rubric evaluations (30 agents $\times$ 15 questions $\times$ 25 months). Agents are rotated monthly by other agents to prevent systematic individual bias and mitigate rater drift over time (Tam *et al*., 2024).

As a single agent evaluates the 15 GPT-generated responses using different conversational evidence within a given month, conventional inter-rater reliability statistics (e.g. Cohen's κ, Fleiss' κ, Krippendorff's α, or intraclass correlation coefficients) are not statistically defined in this setting. Such measures require multiple independent ratings of the same item given the same information. Accordingly, the study does not report formal inter-rater reliability coefficients. Instead, reliability is addressed procedurally through rubric standardization, evaluator training, agent rotation, and large-scale repeated evaluation across time. The rubric's explicit criteria reduce interpretive ambiguity, while the use of trained domain experts ensures informed and contextually grounded judgments. Importantly, the scale and temporal replication of the validation process provide robustness through aggregation: individual judgment variability is averaged out across thousands of evaluations conducted by different agents over two years.

Accordingly, the reported rubric-based accuracy scores should be interpreted as evidence of procedural reliability and consistency across repeated evaluation periods, rather than as population-level estimates of statistical validity. The design supports robustness through standardized expert assessment over time, but it does not permit formal inference about population-level agreement parameters.

The overall accuracy score for a given question is calculated as the average of the scores assigned by the agents in the corresponding month. Figure 6 presents the results of rubric-based validation across the three question types. The average scores fluctuate between 4.4 and 5, tending towards the upper bound, which indicates a high level of accuracy. These findings validate artificial intelligence outputs through natural intelligence.

4.3.2 Artificial intelligence validation. The validation described in the previous section demonstrates a high level of accuracy with the selected sample. However, in practice, within the chosen institution, the model processes an average of 400,000 conversations between agents and customers per execution. It is neither feasible nor practical for a group of individuals to analyze a representative sample of such a large volume of texts. For this reason, the second validation implemented incorporates alternative artificial intelligence algorithms to verify the results produced by GPT-4o Mini. The alternative AI models selected for this purpose are Gemini 2.5 Flash and DeepSeek-V3. These validations are used only for consistency checks, while human rubric validation remains the primary validation mechanism. A comparable validation approach is reported by Laskar *et al*. (2023).

The Jaccard similarity metric is used to assess the degree of textual overlap between the outputs of GPT, Gemini, and DeepSeek. It provides a principled and interpretable method for assessing text similarity when documents are represented as sets of tokens, key phrases, or *k*-shingles (Bag *et al*., 2019). The measure is invariant to document length, which supports its use in near-duplicate detection, plagiarism analysis, and other tasks where the presence or absence of terms is more informative than frequency (Amirzhanov *et al*., 2025; Athukorala and Mohotti, 2022). Compared with alternative measures such as the Szymkiewicz–Simpson coefficient, Euclidean distance, Kullback–Leibler (KL) divergence, and Jensen–Shannon (JS) divergence, Jaccard offers several advantages. It provides a conservative, length-normalized estimate of lexical overlap, thereby enhancing robustness to sparse, noisy short texts (Jiang and Cai, 2024). It also avoids the instability of probabilistic divergences, such as KL and JS, when zero values are present. In addition, it is computationally more efficient than edit-based measures while maintaining resilience to reordering through shingling (Greenberg *et al*., 2023). In addition, its metric properties enable efficient sublinear retrieval using MinHash and

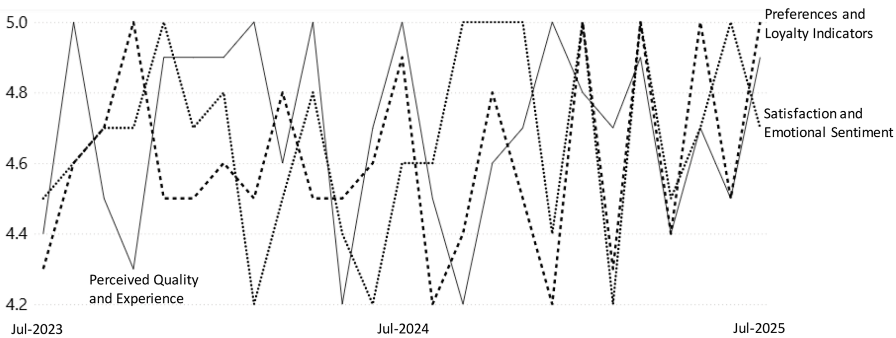

Figure 6. Rubric-based validation for July 2023–July 2025. Source(s): Authors' own work

locality-sensitive hashing. Recent advances have further improved sensitivity and scalability in large collections (Kucherov and Skiena, 2023; Liu *et al*., 2024).

Gemini and DeepSeek serve as independent validators and ground truth to assess GPT's accuracy. In this case, the entire conversation is processed by the three AI APIs. Jaccard similarity quantifies the proportion of shared elements (words, entities, and concepts) between the responses generated by both methods, as expressed in Equation (14).

$$Jaccard(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{14}$$

Where:

A = set of unique items from GPT output and B = set of unique items from Gemini/ DeepSeek output

Example:

GPT: Product credit card is the most mentioned. Customers complain about lost cards.

Gemini: Product credit card is most frequent. Clients complain about lost cards.

Token sets:

GPT set = {product, credit, card, most, mentioned, customers, complain, about, lost, cards}

Gemini set = {product, credit, card, most, frequent, clients, complain, about, lost, cards}

To improve the accuracy of the Jaccard score, the words in the texts being compared are subjected to normalization. Specifically, "lemmatization" converts verbs to their base form, "stopwords" such as "a," "an," "the," and others are removed, and all "synonyms" are represented by a single canonical word using WordNet (via NLTK). With this word normalization, the sets to be compared are defined as follows.

GPT set = {product, credit, card, most, mention, customer, complain, about, lose, card}
Gemini set = {product, credit, card, most, frequent, customer, complain, about, lose, card}
Calculating the union and intersection.

$A \cap B$ = {product, credit, card, most, complain, about, lose, customer, card}

$A \cup B$ = {product, credit, card, most, complain, mention, frequent, about, lose, customer, card}

Thus, the Jaccard score is shown in Equation (15).

$$\text{Jaccard} = 9/11 = 0.8181\ (81.81\%\ \text{overlap}) \tag{15}$$

In addition to the above normalization procedures, to ensure a fair comparison across generative models, a highly constrained prompting strategy is employed. This strategy enforces a fixed output schema, a controlled vocabulary, and evidence-grounded responses. Such a design minimizes linguistic variability and stylistic noise. As a result, high Jaccard values in Figure 7 should therefore be interpreted as evidence of cross-model agreement under constrained prompting and controlled output formatting, rather than as proof of full semantic equivalence between models. In this setting, the metric indicates convergence on extracted customer-experience signals within a tightly specified response regime.

The 30 agents who conduct the rubric-based validation then pose the 15 rubric-defined business-critical questions to the models trained with GPT, Gemini, and DeepSeek. In addition, they formulate and submit 30 open-ended questions, with each agent independently generating one question. The Jaccard scores for the different comparisons are presented in Figure 7.

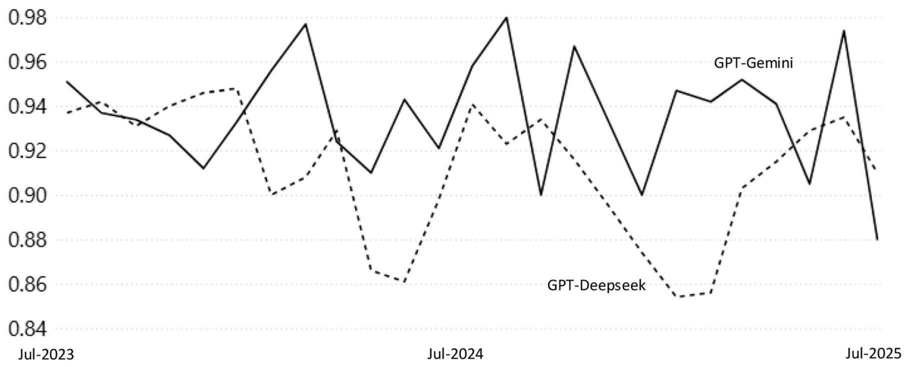

Figure 7. Jaccard score for July 2023–July 2025. Source(s): Authors' own work

In general terms, the proportion of overlap between the results of GPT and Gemini is greater than that between GPT and DeepSeek for the analysis period. However, the Jaccard score exceeds 0.85 with DeepSeek and 0.9 with Gemini in most months, which is considered acceptable for production implementation.

The final product is intended for internal use, specifically within the customer service department. The team will have access to an interface that lets them ask targeted questions about the customer experience in natural language. They will also receive responses in natural language. This functionality enables the extraction of up-to-date insights and knowledge to design mitigating measures that enhance client self-sufficiency and improve the overall experience. As a result, the number of calls requiring agent intervention is minimized. Figures 8 and 9 illustrate a simulation of the current interface, as the internally used UX/UI solutions cannot be displayed due to privacy policies. However, the questions and answers shown correspond to real executions. It represents an interactive system in which the customer experience specialist poses a question, and the bot responds based on conversations

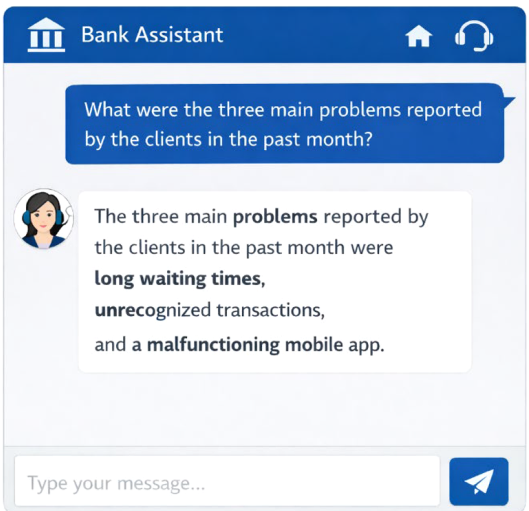

Figure 8. User interface. Note: Question and answer based on conversations from July 2023. Source(s): Authors' own work

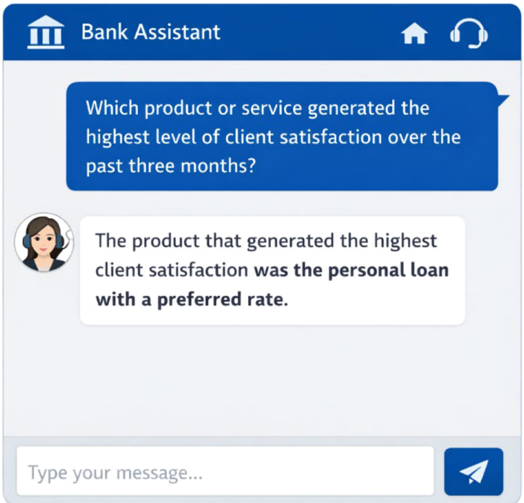

Figure 9. User interface. Note: Question and answer based on conversations from July 2023. Source(s): Authors' own work

received from the contact center. The bot consistently provides real-time, natural language customer insights.

The model proposed in this study has only recently been implemented, and while insights are being generated, the adoption of mitigating measures is still underway. Implementation requires strategic planning, financial resource allocation, and institutional approvals. These processes can be protracted, particularly when they involve modifications to mobile applications. Once strategies are finalized and deployed, the expected outcome is a reduction in complaints, reflected in a smaller set of main topics, even amid continued expansion of products and services.

5. Conclusions

This study demonstrates that integrating competitive learning with generative artificial intelligence (GenAI) offers a robust, scalable, and theoretically grounded approach to extracting customer experience (CX) insights from large-scale conversational data in banking. By analyzing more than 10 million customer–agent interactions, the proposed framework demonstrates how unstructured service conversations can be transformed into actionable, context-rich knowledge. This knowledge surpasses the explanatory power of conventional customer experience instruments such as surveys and Net Promoter Score (NPS), as well as machine learning approaches with structured outputs. Rather than capturing isolated post hoc evaluations, the approach reveals how customer experience is enacted and interpreted during real service encounters, preserving the language, intent, and situational context through which customers articulate value.

From a service-dominant logic (S-D logic) perspective, the proposed method views value in banking as emerging through use and interaction rather than being embedded in products or service outputs. Customer–agent conversations are treated as micro-level episodes of value co-creation. In these episodes, customers integrate institutional resources with their own goals, expectations, and emotions. The framework operationalizes this perspective by identifying recurring interactional contexts where value co-creation succeeds or breaks down.

The study also aligns with relationship marketing theory by treating service interactions as a source of accumulated relational assets, such as trust, perceived competence, and relationship commitment. The generative CX extraction approach captures these relational signals directly from conversational evidence, enabling banks to observe how patterns of responsiveness, empathy, clarity, and problem resolution influence the ongoing quality of customer long-term relationships.

Methodologically, the study advances CX research by demonstrating that competitive learning combined with GenAI can achieve the interpretive richness of human face-to-face interviews at scale. Rubric-based human evaluation and AI benchmarking confirm the reliability, faithfulness, and managerial relevance of the generated insights. Besides, the exclusive use of natural language for both input and output enhances transparency and accessibility, enabling CX professionals to engage directly with the results.

From a managerial standpoint, the findings highlight conversational data as a strategic resource for customer-centric value creation. By continuously monitoring how customers experience services during actual interactions, banks can identify emerging friction points, anticipate dissatisfaction, and design targeted interventions that improve both operational efficiency and relationship quality. Moreover, the ability to pinpoint recurring issues that drive contact center demand provides a foundation for redesigning digital services and self-service functionalities. Such initiatives not only reduce operational costs but also support customer empowerment, a core principle of both S-D logic and relationship marketing, by enabling customers to achieve their goals more independently while strengthening their relationship with the institution.

Despite these contributions, the study has limitations that suggest avenues for future research. The analysis focuses on telephone conversations, excluding other touchpoints such as mobile applications, chat interactions, and in-branch encounters that also contribute to the overall service ecosystem. While rigorous anonymization procedures were applied, ethical and regulatory considerations surrounding conversational data analytics remain an ongoing challenge. In addition, human validation was necessarily conducted on a limited sample of conversations, and AI-based benchmarking relies on current-generation models that may introduce shared biases. Future research should therefore extend the framework across multiple service channels, institutions, and cultural contexts to enhance external validity. Longitudinal studies examining how changes in service design or policy interventions reshape interactional patterns over time would further deepen understanding of value co-creation and relationship development in banking.

References

10 x Banking (2023), “Banks are losing 20% of customers due to poor customer experience”, available at: https://www.10xbanking.com/news/banking-customer-experience-trends (accessed 19 October 2025).

Abdullin, Y., Molla-Aliod, D., Ofoghi, B., Yearwood, J. and Li, Q. (2024), “Synthetic dialogue dataset generation using LLM agents”, *arXiv preprint* arXiv:2401.17461, doi: 10.48550/arXiv.2401.17461.

Adam, M., Wessel, M. and Benlian, A. (2021), “AI-based chatbots in customer service and their effects on user compliance”, *Electronic Markets*, Vol. 31 No. 2, pp. 427-445, doi: 10.1007/s12525-020-00414-7.

Adeniran, I.A., Abhulimen, A.O., Obiki-Osafiele, A.N., Osundare, O.S., Agu, E.E. and Efunniyi, C.P. (2024), “Data-driven approaches to improve customer experience in banking: techniques and outcomes”, *International Journal of Management and Entrepreneurship Research*, Vol. 6 No. 8, pp. 2797-2818, doi: 10.51594/ijmer.v6i8.1467.

Akter, S., Mahmud, F., Rahman, T., Ahmmed, M.J., Uddin, M.K., Alam, M.I., Bhattacharjee, B., Akter, S., Hossain, M.S. and Jui, A.H. (2024), “A comprehensive study of machine learning approaches for customer sentiment analysis in banking sector", *The American Journal of Engineering and Technology*, Vol. 6 No. 10, pp. 100-111, doi: 10.37547/tajet/Volume06Issue10-11.

Amirzhanov, A., Turan, C. and Makhmutova, A. (2025), "Plagiarism types and detection methods: a systematic survey of algorithms in text analysis", *Frontiers of Computer Science*, Vol. 7, 1504725, doi: 10.3389/fcomp.2025.1504725.

An, J. (2026), "Can artificial intelligence enhance service quality? Evidence from US financial services", *International Journal of Bank Marketing*, Vol. ahead-of-print No. ahead-of-print, pp. 1-19, doi: 10.1108/IJBM-04-2025-0300.

Andrian, B., Simanungkalit, T., Budi, I. and Wicaksono, A.F. (2022), "Sentiment analysis on customer satisfaction of digital banking in Indonesia", *International Journal of Advanced Computer Science and Applications*, Vol. 13 No. 3, pp. 466-473, doi: 10.14569/IJACSA.2022.0130356.

Athukorala, S. and Mohotti, W. (2022), "An effective short-text topic modelling with neighbourhood assistance-driven NMF in Twitter", *Social Network Analysis and Mining*, Vol. 12 No. 89, pp. 1-15, doi: 10.1007/s13278-022-00898-5.

Bag, S., Kumar, S.K. and Tiwari, M.K. (2019), "An efficient recommendation generation using relevant jaccard similarity", *Information Sciences*, Vol. 483, C, pp. 53-64, doi: 10.1016/j.ins.2019.01.023.

Bhatnagr, P. and Rajesh, A. (2025), "Online customer experience in Indian digital banks impacting continuous intention usage: generation Y and Z perspective", *Journal of Financial Reporting and Accounting*, Vol. 23 No. 2, pp. 830-850, doi: 10.1108/JFRA-11-2023-0638.

Bhuiyan, R.J., Akter, S., Uddin, A., Shak, M.S., Islam, M.R., Mollick, M.R.A., Rishad, S.M.S.I., Sultana, F. and Hasan-Or-Rashid, M. (2024), "Sentiment analysis of customer feedback in the banking sector: a comparative study of machine learning models", *The American Journal of Engineering and Technology*, Vol. 6 No. 10, pp. 54-66, doi: 10.37547/tajet/Volume06Issue10-07.

Birkenmaier, J. and Stratman, H. (2025), "Consumers' basic bank account complaints and their financial hardships: a content analysis of complaints filed with the consumer financial protection bureau (CFPB)", *Financial Services Review*, Vol. 33 No. 2, pp. 15-35, doi: 10.61190/fsr.v33i2.4057.

Buhler, R.N., De Oliveira Santini, F., Junior Ladeira, W., Rasul, T., Perin, M.G. and Kumar, S. (2024), "Customer loyalty in the banking sector: a meta-analytic study", *International Journal of Bank Marketing*, Vol. 42 No. 3, pp. 513-535, doi: 10.1108/IJBM-08-2023-0484.

Çallı, L. (2023), "Exploring mobile banking adoption and service quality features through user-generated content: the application of a topic modeling approach to Google Play Store reviews", *International Journal of Bank Marketing*, Vol. 41 No. 2, pp. 428-454, doi: 10.1108/IJBM-08-2022-0351.

Chatterjee, S. (2019), "Explaining customer ratings and recommendations by combining qualitative and quantitative user-generated contents", *Decision Support Systems*, Vol. 119, pp. 14-22, doi: 10.1016/j.dss.2019.02.008.

Chen, W., Li, A., Pan, C., Yu, T., Ali, A. and Sun, Y. (2026), "Algorithmic management in the workplace: a systematic review and topic modeling integration using BERTopic", *International Journal of Information Management*, Vol. 86, 102994, doi: 10.1016/j.ijinfomgt.2025.102994.

Défossez, A., Bottou, L., Bach, F. and Usunier, N. (2022), "A simple convergence proof of adam and adagrad", *Transactions on Machine Learning Research*, October, pp. 1-22, available at: https://openreview.net/forum?id=ZPQhzTSWA7

Du, K.-L. (2010), "Clustering: a neural network approach", *Neural Networks*, Vol. 23 No. 1, pp. 89-107, doi: 10.1016/j.neunet.2009.08.007.

Eren, B.A. (2021), "Determinants of customer satisfaction in chatbot use: evidence from a banking application in Turkey", *International Journal of Bank Marketing*, Vol. 39 No. 2, pp. 294-311, doi: 10.1108/IJBM-02-2020-0056.

Eustaquio-Jiménez, R., Durand-Azurza, M., Gamboa-Cruzado, J., León Morales, M., Pajares Ruiz, N., López de Montoya, R. and Ortiz Céliz, N.M. (2024), "Chatbots for customer service in financial entities—a comprehensive systematic review", *Journal of Infrastructure, Policy and Development*, Vol. 8 No. 16, 10122, doi: 10.24294/jipd10122.

Fair Isaac Corporation (FICO) (2025), "New FICO survey: 88% of customers consider customer experience of equal importance to product offerings when choosing their bank", available at: https://www.fico.com/en/newsroom/new-fico-survey-88-customers-consider-customer-experience-equal-importance-product-offerings-when-choosing-their-bank (accessed 19 October 2025).

Fan, Z., He, J., Li, X., Kuang, S., Song, K., Zhou, Y. and Qiu, X. (2025), "CAMIEval: enhancing NLG evaluation through multidimensional comparative instruction-following analysis", *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2025), Volume 1: Long Papers*, Association for Computational Linguistics, pp. 8708-8733, available at: https://aclanthology.org/2025.naacl-long.438/

Farea, A., Tripathi, S., Glazko, G. and Emmert-Streib, F. (2024), "Investigating the optimal number of topics by advanced text-mining techniques: sustainable energy research", *Engineering Applications of Artificial Intelligence*, Vol. 136, 108877, doi: 10.1016/j.engappai.2024.108877.

Fieberg, C., Hornuf, L. and Streich, D.J. (2023), "Using GPT-4 for financial advice: a benchmark study with investor preferences", SSRN Working Paper, doi: 10.2139/ssrn.4499485.

García-Morato, E., Algar, M.J., Alfaro, C., Ortega, F., Gomez, J. and Moguerza, J.M. (2025), "A parametrizable algorithm for distributed approximate similarity search with arbitrary distances", *arXiv preprint* arXiv:2405.13795.

Ghadiridehkordi, A., Shao, J., Boojihawon, R., Wang, Q. and Li, H. (2025), "Leveraging sentiment analysis via text mining to improve customer satisfaction in UK banks", *International Journal of Bank Marketing*, Vol. 43 No. 4, pp. 780-802, doi: 10.1108/IJBM-05-2024-0323.

Greenberg, G., Ravi, A.N. and Shomorony, I. (2023), "LexicHash: sequence similarity estimation via lexicographic comparison of hashes", *Bioinformatics*, Vol. 39 No. 11, pp. 1-9, doi: 10.1093/bioinformatics/btad652.

Hamdan, R.K. (2025), *Sustainable Data Management: Navigating Big Data, Communication Technology, and Business Digital Leadership*, Vol. 1, Springer Cham, Cham, doi: 10.1007/978-3-031-83911-5.

Hashemi, H., Eisner, J., Rosset, C., Van Durme, B. and Kedzie, C. (2024), "LLM-RUBRIC: a multidimensional, calibrated approach to automated evaluation of natural language texts", *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Association for Computational Linguistics, pp. 13806-13834, available at: https://aclanthology.org/2024.acl-long.745/

Hentzen, J.K., Hoffmann, A.O.I., Dolan, R. and Pala, E. (2022), "Artificial intelligence in customer-facing financial services: a systematic literature review and agenda for future research", *International Journal of Bank Marketing*, Vol. 40 No. 6, pp. 1299-1336, doi: 10.1108/IJBM-09-2021-0417.

Hettiarachchi, I. (2025), "The rise of generative AI agents in finance: operational disruption and strategic evolution", *International Journal of Engineering Technology Research and Management*, Vol. 9 No. 4, pp. 447-457, doi: 10.5281/zenodo.15285625.

Hijazi, R. (2022), "Mobile banking service quality and customer value co-creation intention: a moderated mediated model", *International Journal of Bank Marketing*, Vol. 40 No. 7, pp. 1501-1525, doi: 10.1108/IJBM-01-2022-0004.

Hoang, D.P. (2024), "Customer experience of video teller machine service: drivers and behavioural outcomes", *International Journal of Bank Marketing*, Vol. 42 No. 7, pp. 1634-1665, doi: 10.1108/IJBM-09-2023-0528.

Ismael, O. (2025), "The standard deviation score: a novel similarity metric for data analysis", *Journal of Big Data*, Vol. 12 No. 58, pp. 1-42, doi: 10.1186/s40537-025-01091-z.

Jiang, P. and Cai, X. (2024), "A survey of text-matching techniques", *Information*, Vol. 15 No. 6, 332, pp. 1-53, doi: 10.3390/info15060332.

Jim, J.R., Talukder, M.A.R., Malakar, P., Kabir, M.M., Nur, K. and Mridha, M.F. (2024), "Recent advancements and challenges of NLP-based sentiment analysis: a state-of-the-art review", *Natural Language Processing Journal*, Vol. 6, 100059, doi: 10.1016/j.nlp.2024.100059.

Kaur, N. (2021), "Sentiment analysis of e-banking customer reviews using NLP", *ShodhKosh: Journal of Visual and Performing Arts*, Vol. 2 No. 2, pp. 458-465, doi: 10.29121/shodhkosh.v2.i2.2021.5743.

Komulainen, H. and Makkonen, H. (2018), "Customer experience in omni-channel banking services", *Journal of Financial Services Marketing*, Vol. 23 Nos 3-4, pp. 190-199, doi: 10.1057/s41264-018-0057-6.

Komulainen, H. and Saraniemi, S. (2019), "Customer centricity in mobile banking: a customer experience perspective", *International Journal of Bank Marketing*, Vol. 37 No. 5, pp. 1082-1102, doi: 10.1108/IJBM-11-2017-0245.

Krishnan, K. and Easwarakumar, K.S. (2025), "Research topic detection in scientific articles using a hybrid BERT integrated telescopic vector tree model with emperor penguin enhanced NSGA II optimization", *Scientific Reports*, Vol. 15 No. 1, 37201, doi: 10.1038/s41598-025-21145-9.

Kucherov, G. and Skiena, S. (2023), "Improving the sensitivity of MinHash through hash-value analysis", *Proceedings of the 34th Annual Symposium on Combinatorial Pattern Matching (CPM 2023), Leibniz International Proceedings in Informatics (LIPIcs)*, Dagstuhl, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, pp. 20:1-20:12, 20, doi: 10.4230/LIPIcs.CPM.2023.20.

Kumar, P., Mokha, A.K. and Pattnaik, S.C. (2022), "Electronic customer relationship management (E-CRM), customer experience and customer satisfaction: evidence from the banking industry", *Benchmarking: An International Journal*, Vol. 29 No. 2, pp. 551-572, doi: 10.1108/BIJ-10-2020-0528.

Landolsi, H., Letaief, K., Taghouti, N. and Abdeljaoued-Tej, I. (2025), "CAPRAG: a large language model solution for customer service and automatic reporting using vector and graph retrieval-augmented generation", *arXiv preprint* arXiv:2501.13993, doi: 10.48550/arXiv.2501.13993.

Laskar, M.T.R., Bari, M.S., Rahman, M., Bhuiyan, M.A.H., Joty, S. and Huang, J.X. (2023), "A systematic study and comprehensive evaluation of ChatGPT on benchmark datasets", *arXiv preprint* arXiv:2305.18486.

Le, X.C. and Nguyen, T.H. (2024), "The effects of chatbot characteristics and customer experience on satisfaction and continuance intention toward banking chatbots: data from Vietnam", *Data in Brief*, Vol. 52, 110025, doi: 10.1016/j.dib.2023.110025.

Leem, B.H. and Eum, S.W. (2021), "Using text mining to measure mobile banking service quality", *Industrial Management and Data Systems*, Vol. 121 No. 5, pp. 993-1007, doi: 10.1108/IMDS-09-2020-0545.

Liu, R., Deng, G. and He, H. (2024), "Generalized Jaccard feature screening for ultra-high dimensional survival data", *AIMS Mathematics*, Vol. 9 No. 10, pp. 27607-27626, doi: 10.3934/math.20241341.

Mahdikhani, M. and Meena, P. (2024), "Metaverse applications and supply chain innovation: insights from text mining", *Journal of Innovation and Knowledge*, Vol. 9 No. 4, 100591, doi: 10.1016/j.jik.2024.100591.

Mahmood, T., Naseem, S., Ashraf, R., Asif, M., Umair, M. and Shah, M. (2023), "Recognizing factors effecting the use of mobile banking apps through sentiment and thematic analysis on user reviews", *Neural Computing and Applications*, Vol. 35 No. 31, pp. 19885-19897, doi: 10.1007/s00521-023-08827-z.

Makudza, F., Jaravaza, D.C., Makandwa, G. and Mukucha, P. (2024), "Disruptive artificial intelligence: augmenting consumer experience through chatbot banking", in *Digital Influence on Consumer Habits: Marketing Challenges and Opportunities*, Emerald Publishing Limited, Leeds, pp. 1-18, doi: 10.1108/978-1-80455-342-820241001.

Manyanga, W., Makanyeza, C. and Muranda, Z. (2022), "The effect of customer experience, customer satisfaction and word of mouth intention on customer loyalty: the moderating role of consumer demographics”, *Cogent Business and Management*, Vol. 9 No. 1, 2082015, doi: 10.1080/23311975.2022.2082015.

Mittal, D. and Agrawal, S.R. (2022), “Determining banking service attributes from online reviews: text mining and sentiment analysis”, *International Journal of Bank Marketing*, Vol. 40 No. 3, pp. 558-577, doi: 10.1108/IJBM-08-2021-0380.

Mogaji, E. and Nguyen, N.P. (2021), “Managers’ understanding of artificial intelligence in relation to marketing financial services: insights from a cross-country study”, *International Journal of Bank Marketing*, Vol. 40 No. 6, pp. 1272-1298, doi: 10.1108/IJBM-09-2021-0440.

Moharrak, M. and Mogaji, E. (2025), “Generative AI in banking: empirical insights on integration, challenges and opportunities in a regulated industry”, *International Journal of Bank Marketing*, Vol. 43 No. 4, pp. 871-896, doi: 10.1108/IJBM-08-2024-0490.

Mohd Amin, M.R., Asbi, A., Sivakumaran, V.M., Kim, J. and Septiarini, E. (2025), “Artificial Intelligence (AI) adoption in marketing strategies: navigating the present and shaping the future business landscape”, *Social Sciences and Humanities Open*, Vol. 12, 102048, doi: 10.1016/j.ssaho.2025.102048.

Moufok, S., Belkadi, K. and Lebbah, F.Z. (2025), “An artificial ant-based approach using polynomial algorithms to tackle the text aspect of clustering web pages”, *Journal of Applied Data Sciences*, Vol. 6 No. 2, pp. 1050-1075, doi: 10.47738/jads.v6i2.546.

Mozumder, M.A.S., Mahmud, F., Shak, M.S., Sultana, N., Rodrigues, G.N. and Rafi, M.A. (2024), “Enhancing customer satisfaction analysis using advanced machine learning techniques in Fintech industry”, *Journal of Computer Science and Technology Studies*, Vol. 6 No. 3, pp. 27-38, doi: 10.32996/jcsts.2024.6.3.4.

Muniswamaiah, M., Agerwala, T. and Tappert, C.C. (2023), “Applications of binary similarity and distance measures”, *Journal of Systemics, Cybernetics and Informatics*, Vol. 21 No. 3, pp. 11-28, doi: 10.48550/arXiv.2307.00411.

Novotná, T. and Libal, T. (2022), “An evaluation of methodologies for legal formalization”, in Calvaresi, D., Najjar, A., Winikoff, M. and Främling, K. (Eds), *Explainable and Transparent AI and Multi-Agent Systems: EXTRAAMAS 2022, Lecture Notes in Computer Science*, Springer, Cham, Vol. 13283, pp. 189-203, doi: 10.1007/978-3-031-15565-9_12.

Pacella, M., Vasco, P., Papadia, G. and Giliberti, V. (2024), “An assessment of digitalization techniques in contact centers and their impact on agent performance and well-being”, *Sustainability*, Vol. 16 No. 2, p. 714, doi: 10.3390/su16020714.

Palmatier, R.W., Kumar, V. and Harmeling, C.M. (2018), *Customer Engagement Marketing*, Palgrave Macmillan Cham, Cham, doi: 10.1007/978-3-319-61985-9.

Park, Y., Kim, J., Jiang, Q. and Kim, K.H. (2024), “Impact of artificial intelligence (AI) chatbot characteristics on customer experience and customer satisfaction”, *Journal of Global Information Technology Management*, Vol. 27 No. 4, pp. 439-457, doi: 10.1080/21639159.2024.2362654.

Peruchini, M., da Silva, G.M. and Teixeira, J.M. (2024), “Between artificial intelligence and customer experience: a literature review on the intersection”, *Discover Artificial Intelligence*, Vol. 4 No. 4, pp. 1-10, doi: 10.1007/s44163-024-00105-8.

Petersson, A.H., Pawar, S. and Fagerstrøm, A. (2023), “Investigating the factors of customer experiences using real-life text-based banking chatbot: a qualitative study in Norway”, *Procedia Computer Science*, Vol. 219, pp. 697-704, doi: 10.1016/j.procs.2023.01.341.

Pfoertsch, W. and Sulaj, K. (2023), “Integrating artificial intelligence with customer experience in banking: an empirical study on how chatbots and virtual assistants enhance empathy”, *Proceedings of the International Conference on Computing, Networking, Telecommunications & Engineering Sciences Applications (CoNTESA)*, IEEE, pp. 69-74, doi: 10.1109/CoNTESA61248.2023.10384979.

Qiu, Y., Chen, P., Lin, Z., Yang, Y., Zeng, L. and Fan, Y. (2020), “Clustering analysis for silent telecom customers based on K-means++”, *Proceedings of the 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, Vol. 1, pp. 1023-1027, doi: 10.1109/ITNEC48623.2020.9084976.

Qureshi, O.A., Wilkins, S. and Iqbal, H. (2023), "Chatbots: can they satisfy customers in the banking sector?", in Al Marri, K., Mir, F.A., David, S.A. and Al-Emran, M. (Eds), *BUiD Doctoral Research Conference 2023*, Springer, Cham, pp. 360-376, doi: 10.1007/978-3-031-56121-4.

Rahman, S.M., Carlson, J., Gudergan, S.P., Wetzels, M. and Grewal, D. (2022), "Perceived omnichannel customer experience (OCX): concept, measurement, and impact", *Journal of Retailing*, Vol. 98 No. 4, pp. 611-632, doi: 10.1016/j.jretai.2022.03.003.

Ranieri, A., Di Bernardo, I. and Mele, C. (2024), "Serving customers through chatbots: positive and negative effects on customer experience", *Journal of Service Theory and Practice*, Vol. 34 No. 2, pp. 191-215, doi: 10.1108/JSTP-01-2023-0015.

Salem, M. (2024), "Use of chatbots and its role in enhancing customer experience: a study of the banking industry", in Ahmad, S.Z. (Ed.), *Digital Technology and Changing Roles in Managerial and Financial Accounting: Theoretical Knowledge and Practical Application*, Emerald Publishing, Bingley, pp. 107-124, doi: 10.1108/S1479-351220240000036010.

Sathiskumar, N.N. and Andersen, P.H. (2025), "Tensions in AI-driven digital servitization in the banking industry: a socio-technical view", *International Journal of Bank Marketing*, pp. 1-35, doi: 10.1108/IJBM-05-2025-0353.

Schwartz, R., Dodge, J., Smith, N.A. and Etzioni, O. (2020), "Green AI", *Communications of the ACM*, Vol. 63 No. 12, pp. 54-63, doi: 10.1145/3381831.

Sedunov, J. (2020), "Small banks and consumer satisfaction", *Journal of Corporate Finance*, Vol. 60, 101517, doi: 10.1016/j.jcorpfin.2019.101517.

Shamsuzzoha, A., Viitanen, J., Piya, S., Khadem, M. and Shamsuzzaman, M. (2025), "Forecasting financial services of a call center's inbound volumes: a case study", *International Journal of Engineering Business Management*, Vol. 17, pp. 1-14, doi: 10.1177/18479790251367821.

Shreyashree, S., Sunagar, P., Rajarajeswari, S. and Kanavalli, A. (2022), "A literature review on bidirectional encoder representations from transformers", *Inventive Computation and Information Technologies: Proceedings of ICICIT 2021*, Springer, pp. 305-320, doi: 10.1007/978-981-16-6723-7_23.

Steidl, M., Felderer, M. and Ramler, R. (2023), "The pipeline for the continuous development of artificial intelligence models—Current state of research and practice", *Journal of Systems and Software*, Vol. 199, 111615, doi: 10.1016/j.jss.2023.111615.

Tam, T.Y.C., Sivarajkumar, S., Kapoor, S., Stolyar, A.V., Polanska, K., McCarthy, K.R., Osterhoudt, H., Wu, X., Visweswaran, S., Fu, S., Mathur, P., Cacciamani, G.E., Sun, C., Peng, Y. and Wang, Y. (2024), "A framework for human evaluation of large language models in healthcare derived from literature review", *npj Digital Medicine*, Vol. 7 No. 1, pp. 1-14, doi: 10.1038/s41746-024-01258-7.

Tian, Y., Zhang, Y. and Zhang, H. (2023), "Recent advances in stochastic gradient descent in deep learning", *Mathematics*, Vol. 11 No. 3, p. 682, doi: 10.3390/math11030682.

Trivedi, J. (2019), "Examining the customer experience of using banking chatbots and its impact on brand love: the moderating role of perceived risk", *Journal of Internet Commerce*, Vol. 18 No. 1, pp. 91-111, doi: 10.1080/15332861.2019.1567188.

Vargo, S.L. and Lusch, R.F. (2016), "Institutions and axioms: an extension and update of service-dominant logic", *Journal of the Academy of Marketing Science*, Vol. 44 No. 1, pp. 5-23, doi: 10.1007/s11747-015-0456-3.

Vargo, S.L. and Lusch, R.F. (2017), "Service-dominant logic 2025", *International Journal of Research in Marketing*, Vol. 34 No. 1, pp. 46-67, doi: 10.1016/j.ijresmar.2016.11.001.

Verhoef, P.C., Broekhuizen, T., Bart, Y., Bhattacharya, A., Dong, J.Q., Fabian, N. and Haenlein, M. (2021), "Digital transformation: a multidisciplinary reflection and research agenda", *Journal of Business Research*, Vol. 122, pp. 889-901, doi: 10.1016/j.jbusres.2019.09.022.

Voorhees, C.M., Fombelle, P.W., Gregoire, Y., Bone, S., Gustafsson, A., Sousa, R. and Walkowiak, T. (2017), "Service encounters, experiences and the customer journey: defining the field and a call to expand our lens", *Journal of Business Research*, Vol. 79, pp. 269-280, doi: 10.1016/j.jbusres.2017.04.014.

Yeo, W.J., Van Der Heever, W., Mao, R., Cambria, E., Satapathy, R. and Mengaldo, G. (2025), “A comprehensive review on financial explainable AI”, *Artificial Intelligence Review*, Vol. 58 No. 6, p. 189, doi: 10.1007/s10462-024-11077-7.

Zhang, Z., Li, B. and Liu, L. (2023), “The impact of AI-based conversational agent on the firms' operational performance: empirical evidence from a call center”, *Applied Artificial Intelligence*, Vol. 37 No. 1, 2157592, doi: 10.1080/08839514.2022.2157592.

Corresponding author
José-de-Jesús Rocha-Salazar be contacted at: riemmanepsilon@gmail.com