Research on the Mining of Failure Causes of Intelligent Customer Service Based on LDA Topic Model

Shanshan Fang[1,*], Jiwei Yang[1], Zimeng Gui[2], Hongyu Mo[1]
[1] *School of Business, Guilin University of Electronic Technology, Guilin, Guangxi, China*
[2] *School of Mechatronic Engineering, Guilin University of Electronic Technology, Guilin, Guangxi, China*
**Corresponding Author*

Abstract: While the widespread adoption of intelligent customer service systems has significantly boosted service efficiency, the persistent surge in complaints has revealed notable service failure challenges. This study addresses the shortcomings of current analytical approaches, which suffer from inefficiency and high subjectivity. To surmount these obstacles, we utilized the Latent Dirichlet Allocation (LDA) topic model to perform a thorough analysis of 1,247 genuine complaint texts. By means of dual assessment using the perplexity metric and coherence score, we determined that the optimal number of topics is three. We pinpointed three primary issues: functional failure, procedural failure, and emotional failure, which account for 42%, 35%, and 23% of the total, respectively. Our research uncovers that the root causes of intelligent customer service failures lie in the discrepancy between technological capabilities and user expectations, the clash between cost-cutting and efficiency-enhancement measures versus service quality, as well as a bias in system design that favors efficiency over a user-centric approach. Drawing on these insights, this paper puts forward tailored improvement strategies, including enriching the knowledge base, streamlining the process for escalating to human support, and integrating sentiment analysis capabilities. These strategies aim to furnish enterprises with a theoretical foundation and practical guidance for refining their intelligent customer service systems.

Keywords: Service Failure； Text Analysis; Intelligent Customer Service; LDA Topic Model; Topic Mining

1. Introduction

Intelligent customer service systems have emerged as a vital component of enterprise customer service, thanks to their round-the-clock availability, swift response times, and cost-effectiveness. Nevertheless, data sourced from the Black Cat Complaint Platform reveals a nearly 150% surge in complaints related to intelligent customer service over the past three years. These grievances encompass problems like system misinterpretations, challenges in escalating to human assistance, and responses that come across as robotic and unsympathetic, all of which jeopardize user satisfaction and tarnish the reputation of businesses.

Current enterprises predominantly rely on two methods to analyze intelligent customer service failures: manual case sampling or keyword frequency tracking. While manual review is labor-intensive and prone to human bias, keyword-based approaches struggle to contextualize semantic relationships. When dealing with massive volumes of unstructured user feedback, these conventional techniques prove doubly limited—they neither systematically categorize failure patterns nor uncover deeper structural issues. This analytical gap ultimately hinders the design of precision-tuned improvement strategies.

The limitations of these conventional techniques underscore the necessity for more sophisticated analytical approaches.

This need is further amplified by insights from recent studies on AI-mediated service interactions. For instance, research has identified that users' frustration with intelligent customer service often stems from a lack of empathetic response and contextual understanding [1]. The critical role of seamless human-agent handover in mitigating user dissatisfaction has also been emphasized [2]. Moreover, the application of topic modeling to analyze service failures in fintech chatbots has revealed similar functional and emotional gaps [3]. Emerging research is beginning to document the novel challenges posed by generative AI in customer service, such as the generation of factually incorrect or "hallucinated" responses [4]. Lastly, a recent systematic review has called for more empirical studies using real-world complaint data to bridge the gap between AI capabilities and user expectations [5].

This study applies the Latent Dirichlet Allocation (LDA) topic modeling approach to analyze 1,247 authentic customer complaint texts. By eliminating the need for predefined classification frameworks, this data-driven methodology autonomously identifies latent thematic patterns while addressing the inherent limitations of manual analysis in terms of efficiency constraints and subjective bias.

The analytical framework thus established provides empirical support for enterprises to strategically enhance their intelligent customer service systems. The research comprises four key components: systematic data acquisition and preprocessing, development and refinement of the LDA model, comprehensive analysis and causal attribution of emergent themes, and formulation of actionable managerial recommendations. Aligned with the stated research objectives and methodology, the paper's structure is organized as follows: Section 2 conducts a critical review of scholarly works pertaining to intelligent customer service systems, service failure mechanisms, and text mining techniques.

Section 3 details the methodological framework, including data provenance, preprocessing protocols, and LDA model implementation. Section 4 interprets the empirical findings derived from the topic modeling analysis. Section 5 examines the theoretical contributions and practical implications of the study's outcomes. The final section concludes with a synthesis of key insights and outlines potential avenues for future research.

2. Literature Review

2.1 Research Progress on Intelligent Customer Service Systems

Intelligent customer service (ICS) leverages natural language processing (NLP) and machine learning (ML) technologies to achieve service automation. Early research primarily emphasized technical viability and operational efficiency gains, while recent studies have increasingly scrutinized systemic limitations. According to Huang and Rust's hierarchy of service intelligence, intelligent systems excel in mechanical and analytical intelligence domains but demonstrate significant deficiencies in emotional intelligence capabilities [6].Empirical evidence indicates a persistent user preference for human agents when handling complex or emotionally charged service scenarios [7]. Gnewuch et al. identified persistent gaps in conversational AI's intent recognition accuracy [8], while Luo et al. documented operational challenges arising from enterprises' over-reliance on automated systems, particularly regarding seamless human agent handover [9]. Mende et al. further corroborated that ICS platforms lack fundamental empathetic capacities [10].However, methodological limitations persist in existing literature. The majority of studies rely on controlled laboratory experiments or qualitative interviews with limited sample sizes, resulting in a dearth of large-scale empirical analyses using authentic customer complaint datasets. This research gap underscores the need for systematic investigation of real-world service failure contexts.

2.2 Theoretical Foundations of Service Failure

Service failure denotes instances where service delivery deviates from customer expectations. Bitner et al. [11] conceptualized three primary failure types: core service failures, process failures, and employee behavior failures. Smith et al. [4] further demonstrated that failure severity is moderated by three interrelated factors: the nature of the failure event, customer attribution patterns, and the effectiveness of recovery strategies [12]. In human-computer interaction (HCI) contexts, Parasuraman and Colby revealed fundamental expectation discrepancies between technical systems and human service providers. They posited that failure attribution follows distinct cognitive pathways when users interact with intelligent systems, as technical failures tend to be perceived as systemic rather than intentional. This shifts the blame locus from service personnel to algorithmic limitations or design flaws [13].However, existing theoretical frameworks exhibit critical limitations. The preponderance of service failure research remains grounded in traditional service encounters characterized by direct human-to-human contact. These models inadequately account for the unique failure modalities emerging in AI-mediated service environments, particularly regarding: the dual-process attribution mechanism, the emotional dissonance caused by machine empathy deficits, and the irreparable trust damage resulting from algorithmic opacity.

2.3 Application of Text Mining in the Customer Service Domain

The topic model is capable of automatically uncovering the latent semantic structure embedded within text. The LDA model, introduced by Blei et al. [14] and colleagues, employs Bayesian inference to ascertain the topic distribution across documents, showcasing remarkable proficiency in managing high-dimensional, sparse textual datasets [14]. Guo and his team leveraged LDA to dissect hotel reviews, pinpointing the central themes [15], while He and associates constructed models of telecommunications complaint texts to trace the primary origins of issues [16]. Nevertheless, the utilization of LDA in research concerning the failures of intelligent customer service systems remains an uncharted territory. Moreover, methodological considerations, including the scientific determination of the optimal number of topics and the integration of qualitative validation to guarantee the precision of interpretations, necessitate standardization.

2.4 Research Review

A thorough examination of the existing literature reveals that notable gaps persist in the interdisciplinary convergence among the three domains: intelligent customer service, service failure, and text mining.

Existing studies on intelligent customer service predominantly center on assessing technical performance, with discourse on service failures largely confined to conceptual frameworks rather than empirical investigations grounded in real-world complaint data. Prevailing service failure theories, primarily derived from traditional human-mediated service contexts, lack systematic exploration of the unique failure mechanisms inherent to AI-driven service providers. While topic modeling has gained traction in general customer feedback analysis, its application to intelligent customer service complaint texts remains virtually unexplored. Critical methodological challenges-such as the scientific determination of optimal topic quantities and the integration of qualitative validation to ensure thematic interpretability-lack of standardized protocols. Furthermore, current research stagnates at the identification of failure patterns, with insufficient scrutiny into their systemic root causes and a notable absence of structured pathways to translate findings into actionable service optimization strategies.

The study's originality is threefold. First, in terms of research focus, it centers on intelligent customer service—a paradigmatic human-AI interaction context—to systematically dissect the typology and causal mechanisms of service failures, addressing a theoretical void in technology-driven service domains. Second, from a methodological standpoint, the study pioneers the application of LDA topic modeling to large-scale real-world complaint datasets. By employing dual validation through perplexity metrics and coherence scoring, it identifies optimal topic structures, while integrating qualitative validation to bolster interpretive rigor. This approach establishes a standardized text-mining framework for intelligent customer service research. Third, regarding practical value, the study transcends mere failure classification by uncovering systemic root causes. It proposes actionable optimization strategies across three dimensions—functional, procedural, and emotional—thereby constructing a holistic research pathway from problem diagnosis to evidence-based strategy development.

3. Research Process

3.1 Data Collection

Following a comparative analysis of multiple data sources, the Black Cat Complaint Platform (a Sina-affiliated consumer advocacy portal) was selected as the primary data repository. Unlike structured enterprise internal records, user-generated complaints on this platform predominantly reflect spontaneous expressions of consumer grievances. This unfiltered nature enables more authentic insights into problem perception and emotional reactions. While the platform does contain emotionally charged narratives, such expressions provide critical contextual clues for deciphering service failure mechanisms.

The data collection window was defined as May to October 2025. While researchers initially contemplated a 12-month observation period, pilot testing revealed that the typology of intelligent customer service complaints reached stability within six months. Prolonging the duration beyond this point risked introducing confounding variables, particularly platform algorithm adjustments. The retrieval process employed "intelligent customer service" as the primary search term, supplemented by synonymous expressions including "robotic customer service" and "automated response systems." This approach yielded an initial dataset comprising approximately 2,000 complaint records.

The preliminary retrieval generated approximately 2,000 records, though not all met the study's inclusion criteria. Specifically, certain complaints referenced intelligent customer service systems but primarily concerned product defects or logistical issues, necessitating their exclusion. Additionally, a subset of records contained insufficient descriptive detail to support thematic analysis, while duplicate submissions were identified through semantic similarity algorithms. Following a dual-stage validation process combining manual review with automated filtering, 1,247 valid cases were retained. While this sample size may appear modest, the research team determined it adequate for topic modeling purposes given the method's emphasis on semantic density over sheer volume.

The narrative content of each complaint constitutes the primary unit of analysis, averaging approximately 180 words in length. Notably, variations in textual volume correlate with the intensity of grievance expression: longer complaints typically manifest heightened dissatisfaction and involve more intricate situational contexts. All personally identifiable information (PII) was systematically redacted during the data acquisition phase, adhering to GDPR-compliant anonymization standards.

3.2 Data Preprocessing

To ensure robust text mining outcomes, we performed systematic preprocessing on the collected complaint dataset. The raw complaint texts contained various noise elements, including order IDs, URLs, and special characters, requiring batch cleaning via regular expressions. Notably, we selectively preserved exclamation marks and question marks during this process, as these punctuation cues effectively capture users' emotional expressions.

Subsequently, we employed the Jieba tokenizer for Chinese text segmentation. To optimize professional term recognition, we augmented the base dictionary with domain-specific terminology including "intelligent customer service", "transfer to human agent", and "refund processing". Following tokenization, we constructed a tailored stop word repository by integrating the Harbin Institute of Technology's stop word list with additional high-frequency yet low-semantic-value terms such as "customer service" and "chatbot", thereby enhancing the quality of textual features for subsequent analysis.

Following the comprehensive preprocessing pipeline—including data cleansing, tokenization, and noise reduction—we constructed a high-quality standardized corpus comprising 1,247 validated complaint documents. The cleansing outcomes are summarized in Table 1, while Figure 1 presents a word cloud visualization highlighting frequently occurring terms, where font size corresponds directly to word frequency.

3.3 LDA Topic Model Construction

3.3.1 Theoretical foundations

Latent Dirichlet Allocation (LDA), introduced by Blei et al. [14], represents a paradigmatic generative probabilistic model for topic discovery. The model posits that each document manifests as a probabilistic mixture of latent topics, where each topic corresponds to a probability distribution over the vocabulary. Through statistical analysis of word co-occurrence patterns, LDA performs reverse inference to uncover both the latent topic composition of documents and the word distributions characterizing each topic.

The selection of LDA in this study is grounded in three key considerations: First, as an unsupervised learning framework, LDA eliminates the need for pre-annotated training data, making it particularly suitable for exploratory research scenarios. Second, the model generates topic-word distributions with high interpretability, enabling clear identification of specific problem categories in intelligent customer service complaints. Third, LDA has demonstrated proven effectiveness in Chinese text topic mining tasks, with established technical methodologies ensuring implementation reliability.

Having established the suitability of the LDA model, the subsequent critical step involves determining the optimal number of topics (k) to ensure the model accurately reflects the underlying structure of the complaint data.

Table 1. Comparison before and after Text Cleaning

Original comment	The result after tokenization cleaning	
I made a 30-minute call to the customer service, but the robot kept asking the same questions repeatedly, and each time it gave the same template responses. I couldn't understand my real needs at all. In the end, the problem still wasn't solved. I was transferred to a human customer service representative and had to wait for 40 minutes before getting through. The whole process was so infuriating. What kind of service is this!	half hour	repeatedly
	matter	template
	reply	cannot understand
	demand	true
	solve	transfer
	put through	artificial
	await	robot

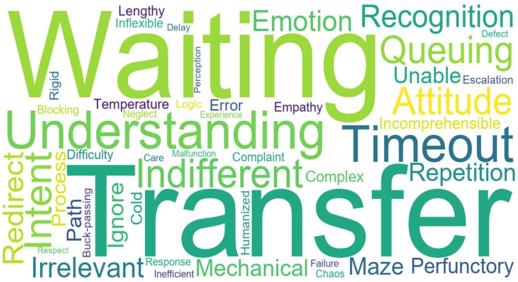
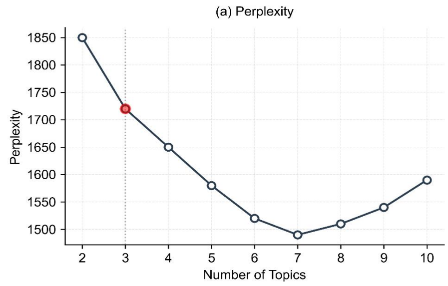


Figure 1. Word Cloud of Failed Intelligent Customer Service Cases

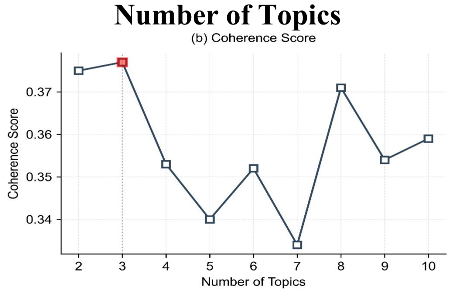
3.3.2 Topic quantity optimization

The configuration of topic count significantly influences the interpretive capacity of LDA models. Excessive topics induce semantic fragmentation, yielding under-representative themes that lack coherent conceptual focus. Conversely, insufficient topics fail to adequately capture data heterogeneity, resulting in overly generalized topic semantics that compromise discriminative power. To address this trade-off, this study employs a hybrid evaluation metric combining perplexity and topic coherence to identify the optimal number of topics.

Based on an analytical corpus of 1,247 complaint texts, this study established a candidate range of 2-10 topics for systematic evaluation. For each candidate topic count, we computed both perplexity and topic coherence, with comparative results visualized in Figure 2. The perplexity trajectory exhibited a pronounced inflection at k=3, where the score plummeted to 1,720 before ascending with increasing topic counts—a pattern indicative of emerging overfitting. Concurrently, the topic coherence metric peaked at 0.377 when k=3, demonstrating statistically significant superiority over alternative configurations. This dual-metric analysis revealed that the k=3 configuration achieved optimal equilibrium between model parsimony and semantic fidelity. The substantial reduction in perplexity coupled with maximal coherence scores confirmed this as the theoretically justified topic count for subsequent analysis.



(a) The Trend of Perplexity Varying with the Number of Topics



(b) The Trend of Consistency Score as the Number of Topics Changes

Figure 2. Determination of the Number of Topics in the LDA Mode

The findings demonstrate strong alignment with operational patterns observed in smart customer service complaint scenarios. Building on established research that categorizes service failures across three core aspects —system functionality, service process, and emotional engagement—our thematic framework not only mirrors this conceptual structure but also significantly improves the model's explanatory power.

3.4 Thematic Result Analysis

3.4.1 High-frequency word analysis

A word frequency analysis was conducted on 1,247 preprocessed complaint texts, yielding 2,865 distinctive terms. Table 2 presents the 24 most frequently occurring keywords alongside their counts. Notably, "processing," "procedure," and "system" emerged as the top three terms with significantly higher occurrence rates than other words. This pattern suggests that user complaints primarily focus on two dimensions:

the efficiency of service handling procedures, and the technical robustness of the system itself. Additional high-frequency terms such as "attitude," "resolution," and "channel" further reveal specific concerns related to rights assertion and interaction channels. Meanwhile, the prevalence of negative terms including "defect," "failure," and "ineffective" directly indicates users' adverse perceptions of service breakdowns.

Table 2. Frequent Keywords for Complaints from Black Cats

subject term	frequency	subject term	frequency
dispose	654	intelligent	162
process	624	ineffective	150
channel	621	play at	148
system	543	deficiency	147
flaw	450	consult with	126
lose efficacy	438	sales return	123
systematicness	393	after sale	111
refund	357	voluntarily	99
merchant	348	buckpassing	87
appeal	195	change-over	84
solve	178	function	81
reply	171	response	82

3.4.2 Text topic analysis

Figure 3 provides a visual representation of the LDA model's output through inter-topic distance mapping. Each bubble in the visualization corresponds to an identified topic, with its area proportional to "relevance score" calculated by pyLDAvis. This metric reflects the topic's prevalence and distinctiveness across the entire corpus. In general, topics represented by larger bubbles demonstrate greater significance or prominence within the document collection. The spatial arrangement of bubbles further reveals semantic relationships between topics: closer proximity indicates stronger conceptual overlap, while greater distances suggest weaker semantic associations.

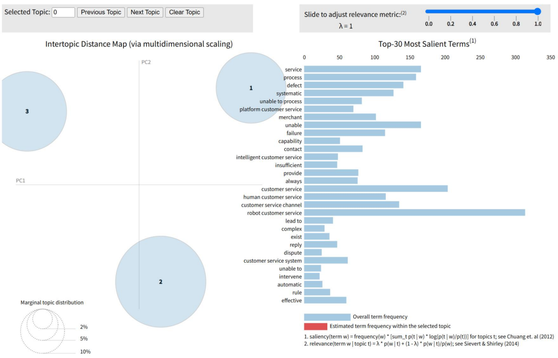


Figure 3. Semantic Bubble Chart of Topic Distribution

As illustrated in Figure 3, the three thematic bubbles exhibit a balanced distribution across distinct quadrants, with minimal overlapping regions. This spatial arrangement indicates that the model demonstrates satisfactory fitting performance under the specified number of topics. Each identified topic not only captures distinct conceptual dimensions within the corpus but also preserves clear semantic boundaries. The thematic analysis yields highly interpretable results with strong discriminative power. Although slight intersections exist between certain bubbles, suggesting shared lexical features among some topics, this phenomenon aligns with the fundamental mechanism of LDA models—which generate latent topics through probabilistic analysis of textual content. Such inter-topic relationships reflect the inherent interconnectedness of service failure causes documented in real-world complaints.

The analysis of vocabulary distribution and weights across the three identified themes is presented in Table 3. Theme 1 constitutes the largest proportion at 43.3%, with core keywords including system, defect, failure, intelligence, error, which collectively highlight fundamental technical issues. Theme 2 accounts for a significant 32.1%, featuring terms such as processing, workflow, channel, response, evasion—indicating bottlenecks in service operation and management processes. Theme 3 represents 24.6% of the distribution, with keywords like refund, merchant, appeal, after-sales, fraud pointing to deficiencies in consumer rights protection mechanisms.

Table 3. Themes - Key Words

theme	subject term		
Technical system defect	system	rigidify	lose efficacy
	flaw	intelligence	automatic
	fault	function	systematicness
	logic	recognition	malfunction
Inefficient service process	dispose	process	channel
	solve	change-over	response
	reply	buckpassing	buckpassing
	play at	deficiency	delay
Lack of rights protection	refund	merchant	compensate
	fraud	sales return	after sale
	false	consult with	dispute
	appeal	order form	can not

The LDA topic model has objectively clustered keywords extracted from complaint texts. To enhance the interpretive power of identified themes, this study follows established scholarly practices by conducting manual coding and thematic naming based on semantic coherence and conceptual linkages among clustered terms within each topic. This process transforms model-generated keyword clusters into first-order themes with explicit semantic definitions, effectively translating abstract statistical outputs into interpretable conceptual frameworks.

Building upon this foundation, the study synthesized three thematic clusters emerging from the LDA analysis, labeling them as Systemic Technical Deficiencies, Service Process Inefficiencies, and Rights Protection Deficits, with detailed mappings presented in Table 3. The analysis reveals a cascading failure mechanism in intelligent customer service systems: breakdowns originate from systemic technical deficiencies as the root cause, propagate through operational inefficiencies in service delivery processes, and ultimately culminate in rights protection deficits that directly trigger user dissatisfaction and complaints. This trajectory clearly delineates the origin-to-impact pathway of service failures, demonstrating how technical malfunctions escalate into procedural failures before manifesting as tangible rights violations.

3.5 Summary of Key Findings

Through probabilistic topic modeling of 1,247 consumer complaint narratives, this study systematically distilled three principal failure pathways from the heterogeneous unstructured corpus.

The functional dimension constitutes roughly 42% of the issue spectrum. Complaint narratives frequently feature expressions like "I don't understand," "irrelevant responses," and "repetitive answers," which collectively underscore the intelligent system's limitations in intent recognition. This phenomenon stems not solely from technical constraints but also from critical knowledge base deficiencies—as evidenced by numerous user queries failing to retrieve precise matches, forcing the system into vague feedback loops.

Process-related barriers constitute roughly 35% of service disruptions. Recurrent complaints—including "can't escalate to human agent," "prolonged hold times," and "endless transfer loops"—expose systemic over-reliance on automation strategies: Cost-containment measures have rendered human escalation pathways excessively convoluted, effectively stranding users in a labyrinthine customer service ecosystem.

The emotional dimension constitutes roughly 23% of user dissatisfaction. Recurrent descriptions like "mechanically indifferent" and "emotionally barren" reveal systemic deficits in affective computing capabilities. During emotional escalation scenarios—where users exhibit heightened agitation—the system fails to activate adaptive escalation protocols, resulting in prolonged service isolation from human intervention.

At a deeper level, these three types of failures reveal structural mismatches between technological capabilities and user experiential expectations. More fundamentally, they expose inherent contradictions between corporate cost-efficiency optimization strategies and service quality preservation imperatives. Ultimately, the systemic design paradigm that prioritizes operational efficiency over user-centeredness constitutes a fundamental conceptual deviation. Without rectifying these systemic flaws, even the most technologically sophisticated solutions will remain incapable of securing sustained user trust.

4. Conclusion

4.1 Theoretical Implications Theoretical Implications

The study's theoretical advancements primarily manifest in three dimensions: contextual enrichment of service failure theory through empirical validation in intelligent service environments; methodological robustness demonstrated via cross-case comparison of failure patterns across digital service ecosystems; and conceptual scaffolding for an integrative framework that systematically maps the causal pathways of intelligent service failures.

Firstly, this study extends the theoretical domain of service failure from conventional interpersonal contexts to algorithm-mediated human-computer interactions, exemplified by intelligent service systems. Such extension substantiates the enduring relevance of classical service failure frameworks in AI-driven environments, while necessitating ontological reconceptualization aligned with algorithmic governance paradigms. Notably, the identified systemic technological deficiencies transcend traditional outcome-centered failure taxonomies, unveiling critical epistemic dimensions encompassing human-AI trust dynamics and algorithmic reliability validation.

Secondly, from a methodological standpoint, this research empirically validates the distinctive hermeneutic capacity of the LDA topic model as an unsupervised machine learning framework for decoding latent semantic structures in unstructured consumer complaint data. In contrast to a priori content analysis, this computational approach enables the emergent identification of core thematic clusters from large-scale datasets, while its algorithmic transparency and procedural replicability establish a methodologically rigorous analytical paradigm for subsequent consumer behavior research—particularly during exploratory phases of inquiry.

Grounded in empirical findings, this investigation establishes a tripartite configuration encapsulating intelligent service system failures along functional, procedural, and affective dimensions. Beyond systematically mapping critical failure manifestations in technologically mediated service encounters, the framework elucidates structural interdependencies between failure modalities, thereby establishing a conceptual scaffolding for future scholarly interrogation of causal mechanisms and cascading effect trajectories.

4.2 Managerial Implications

This study delineates actionable pathways for enterprises to refine their intelligent customer service systems, offering targeted strategic guidance grounded in empirical evidence.

At the functional level, enterprises must acknowledge that knowledge base deficiencies constitute a primary bottleneck constraining the system's contextual comprehension. Models trained exclusively on generic corpora typically exhibit inadequate domain-specific interpretive capacity, necessitating targeted efforts to curate industry-specific "question-answer" exemplar repositories. Crucially, implementing a self-adaptive learning framework enables continuous system refinement through each user interaction—a process characterized not by isolated optimization events, but by sustained evolutionary adaptation.

In process design, numerous enterprises currently impose excessively stringent thresholds for human agent escalation, forcing users through multiple unproductive interactions before accessing live support. From a user experience standpoint, automatic transfer after three consecutive unresolved sessions would align more closely with service expectations. The persistent shortage of human agents during peak demand periods warrants urgent attention, as short-term cost optimization strategies often result in irreparable reputational damage over time. Furthermore, data analysis reveals that overly granular problem categorization systems in certain organizations create latent process friction, requiring users to navigate convoluted option hierarchies to locate relevant solutions—a design flaw that inherently undermines service efficiency.

In the emotional dimension, passive identification of negative emotions falls short. The critical imperative lies in establishing proactive escalation pathways. For users with recurrent complaints, dedicated VIP resolution channels not only elevate satisfaction but also operationalize the enterprise's service ethos. From a strategic vantage point, organizations must re-examine the fundamental role of intelligent customer service: a cost-containment mechanism or an experience-enhancement catalyst? These paradigms entail fundamentally divergent system architectures—the former prioritizes self-service metrics, while the latter emphasizes resolution efficacy. This ideological divergence permeates every operational layer, from knowledge engineering and process orchestration to affective strategy formulation.

4.3 Limitations and Future Directions

This study presents three key limitations. First, data collection was confined to the Black Cat Complaint Platform, with samples primarily from complaining users, thereby excluding the experiences of satisfied or neutral users. Future research should incorporate multi-source data (e.g., social media, enterprise ticketing systems) for cross-validation. Second, the research design prioritized failure cause identification but overlooked differential impacts of failure types on user behavior. Subsequent studies could employ structural equation modeling to quantify causal pathways among variables. Third, while the LDA model was validated with three topics, its static framework may fail to capture nuanced semantic variations. Dynamic topic modeling or deep learning approaches warrant exploration.

Future research in intelligent customer service systems may prioritize three key dimensions. First, longitudinal investigations could systematically examine sustained behavioral changes induced by service failures. Second, cross-sector comparative analyses would clarify industry-specific failure mechanisms. Third, empirical studies on service recovery processes may offer organizations actionable frameworks to rebuild user trust. Concurrently, the proliferation of generative AI necessitates academic scrutiny of emergent failure risks inherent in advanced technological implementations.

Acknowledgments

This paper received the funding from the college Student Innovation and Entrepreneurship project of"A Study on the 'Double-Edged Sword' Effect of AI-Powered E-Commerce Customer Service Usage on Consumer Behavior Based on the SOR Model" (Project number: 202510595035).

References

[1] Li X, Park S. Human-AI interaction in customer service: The role of escalation mechanisms. Computers in Human Behavior. 2023; 142: 107650.

[2] Zhang L, Wang Y, Chen H. Understanding finchatbot failures: A topic modeling approach. International Journal of Bank Marketing. 2023; 41(2): 345-364.

[3] Kumar A, Smith R, Lee J. AI in customer service: A systematic review of failures and recovery strategies. Journal of Retailing and Consumer Services. 2024; 76: 103558.

[4] Smith J, Yang L, Garcia P. The Promise and Peril of Generative AI in Customer Service: Analyzing Hallucinations and Reliability Issues. Journal of Service Management. 2024; 35(3): 245-267.

[5] Chen X, Kim S Y. Building and Breaking Trust: User Experiences with AI Service Failures and the Recovery Paradox. Computers in Human Behavior. 2025; 162: 107952.

[6] Huang M H, Rust R T. Artificial intelligence in service. Journal of Service Research, 2018, 21(2): 155-172.

[7] Larivière B, Bowen D, Andreassen T W, et al. "Service Encounter 2.0": An investigation into the roles of technology, employees and customers. Journal of Business Research, 2017, 79: 238-246.

[8] Gnewuch U, Morana S, Maedche A. Towards designing cooperative and social conversational agents for customer service. Proceedings of the 38th International Conference on Information Systems. Seoul, 2017.

[9] Luo X, Tong S, Fang Z, et al. Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. Marketing Science, 2019, 38(6): 937-947.

[10] Mende M, Scott M L, van Doorn J, et al. Service robots rising: How humanoid robots influence service experiences and elicit compensatory consumer responses. Journal of Marketing Research, 2019, 56(4): 535-556.

[11] Bitner M J, Booms B H, Tetreault M S. The service encounter: diagnosing favorable and unfavorable incidents. Journal of Marketing, 1990, 54(1): 71-84.

[12] Smith A K, Bolton R N, Wagner J. A model of customer satisfaction with service encounters involving failure and recovery. Journal of Marketing Research, 1999, 36(3): 356-372.

[13] Parasuraman A, Colby C L. An updated and streamlined technology readiness index: TRI 2.0. Journal of Service Research, 2015, 18(1): 59-74.

[14] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet allocation. Journal of Machine Learning Research, 2003, 3: 993-1022.

[15] Guo Y, Barnes S J, Jia Q. Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent Dirichlet allocation. Tourism Management, 2017, 59: 467-483.

[16] He W, Zha S, Li L. Social media competitive analysis and text mining: A case study in the pizza industry. International Journal of Information Management, 2013, 33(3): 464-472.