GAS Journal of Economics and Business Management (GASJEBM)
Volume 3 | Issue 8, 2026 Homepage: https://gaspublishers.com/gasjebm-home/ ISSN: 3048-782X

# Intelligent Identification of Consumers' Implicit Needs and Quality Adaptation Strategies Based on Text Mining

**Wenxin Xu, Junchao Yang**

Southwest Petroleum University, Chengdu, 610500, China

***Received:*** *05.07.2026* | ***Accepted:*** *20.08.2026* | ***Published:*** *22.08.2026*

***Corresponding Author:** Wenxin Xu

**DOI:** *10.5281/zenodo.22055016*

**Abstract** **Original Research Article**

In the digital consumption era, consumer feedback texts contain substantial implicit needs that traditional surveys fail to capture due to sample bias and expression barriers. This study, based on 62,716 e-commerce reviews, constructs a five-layer framework—data collection, text mining, implicit need identification, need-quality mapping, and quality adaptation strategy—using jieba segmentation, LDA topic modeling, SnowNLP sentiment analysis, and keyword-pattern matching. Results show that: (1) LDA extracts 8 structured topics from all reviews and 6 pain-point topics from negative reviews; (2) 36,486 implicit need records are identified across 8 dimensions; (3) The SECI model's externalization stage is operationalized through text mining; (4) Kano classification yields three priority levels (P0, P1, P2) with differentiated strategies. This research offers a methodological reference for enterprises to identify consumer needs and optimize quality management in digital transformation.

**Keywords:** Text Mining; Implicit Need Identification; SECI Model; Kano Model; Knowledge Management.



## 1. Introduction

With the rapid development of e-commerce, it has become common for consumers to express their product experiences through online reviews. Data from the China Internet Network Information Center shows that by the end of 2024, the number of online shopping users in China had exceeded 900 million, generating tens of millions of product reviews daily. These review texts not only contain consumers' direct evaluations of product functions but also harbor numerous unexpressed expectations, dissatisfactions, and improvement suggestions—namely, "implicit needs." Unlike "explicit needs" that consumers can clearly articulate, implicit needs are often hidden in vague expressions, emotional complaints, or indirect comparisons, making them difficult to capture effectively with traditional questionnaire methods. Han et al. (2023) utilized deep learning language models to extract latent needs from online reviews, validating the important value of review texts as a source of implicit need information. Wang et al. (2022) further proposed methods for discovering product improvement features from users' implicit feedback, demonstrating that implicit need mining is of critical significance for product innovation.

In the field of quality management, accurately identifying consumer needs is the starting point for quality improvement. The Kano model, proposed by Kano in 1984, classifies consumer needs into three types: basic, performance, and attractive, pointing out that different types of needs have fundamentally different impact mechanisms on satisfaction (Kano et al., 1984). However, traditional need identification methods mainly rely on questionnaires and focus group interviews, which suffer from problems such as insufficient sample representativeness, respondent recall bias, expression barriers, and high costs. Joung and Kim (2022) pointed out that traditional Kano questionnaire methods have issues with response bias and time lag, proposing an automatic Kano classification method based on neural networks from online reviews. The research by Ananthakrishnan et al. (2023) further verified the significant positive impact of consumer voice (online reviews) on enterprise quality improvement. The rise of text mining technology offers new possibilities for breaking through the bottlenecks of traditional methods—through automated analysis of massive unstructured text data, it is possible to discover consumers' real needs on a larger scale and at a deeper level.

In recent years, research on the application of text mining technology in consumer need identification and quality management has grown rapidly, forming a relatively rich body of research around topic modeling (Miao et al., 2023; Li et al., 2020; Zhang et al., 2019), need classification, quality mapping, and intelligent technology applications. Existing studies have confirmed the effectiveness of methods such as LDA topic modeling (Miao et al., 2023), sentiment analysis, and deep learning in the structured mining of user needs, have gradually explored the integration of the Kano model (Dong et al., 2025; Jin et al.,2022; Chen et al., 2022) and QFD (Lee et al., 2024; Qin et al., 2024; Mohseb, A. 2025) with text mining, and have begun to pay attention to the theoretical expansion and practical application of the SECI (Böhm et al., 2026; Ranjan et al., 2026; Wang, J. 2026) knowledge conversion model in digital scenarios. Nevertheless, current research still has obvious shortcomings: most studies remain at the surface level of sentiment analysis and topic mining, failing to systematically deconstruct the implicit need structure behind reviews; there is a lack of systematic integration of text mining results with quality management theories such as Kano and QFD; there is little interpretation of the knowledge conversion mechanism of consumers' implicit needs from the perspective of knowledge management, and no effective connection has been established between SECI theory and user need mining; at the same time, existing research mostly focuses on need identification, lacking a quantitative evaluation system for need intensity and priority, making it difficult to support refined quality decision-making.

Based on the above research status and existing research gaps, this study focuses on the mining of consumers' implicit needs and product quality optimization in the e-commerce review scenario, addressing four core research questions: How to systematically identify consumers' implicit needs from massive e-commerce reviews using text mining techniques? How to transform fragmented, unstructured implicit needs into standardized, actionable product quality attributes? How to explain the complete transformation process of consumers' implicit needs from tacit perception to explicit implementation based on SECI knowledge management theory? How to formulate precise quality adaptation and product improvement strategies based on differentiated need characteristics?

## 2 Literature Review

### 2.1 Research on SECI Knowledge Conversion Theory in Digital Scenarios

The SECI knowledge conversion model is a classic theoretical framework in the field of knowledge management, proposed by Nonaka and Takeuchi (1995), which systematically explains the evolutionary pattern of cyclical conversion and spiral ascent between tacit knowledge and explicit knowledge in the process of organizational knowledge creation, encompassing four core stages: Socialization, Externalization, Combination, and Internalization. Nonaka further clarified the core characteristics of tacit knowledge as personalized, contextual, and difficult to standardize and encode, pointing out that metaphor, analogy, and external empowerment are key pathways for the externalization of tacit knowledge, laying a theoretical foundation for subsequent research on knowledge conversion in digital scenarios (Nonaka, I.1994).

With the rapid development of digital technologies and generative artificial intelligence, the academic community has begun to focus on the digital adaptation and contextual expansion of the SECI model, promoting the iteration of traditional knowledge conversion theory towards an intelligent knowledge management paradigm. Böhm and Durst (2026) proposed a knowledge management evolution framework from SECI to GRAI, systematically demonstrating the automated empowerment value of AI technology in the externalization stage of tacit knowledge, pointing out that the traditional SECI model urgently needs to be updated theoretically to adapt to generative AI scenarios. Ranjan and Joshith (2026) empirically verified the applicability of the SECI model for tacit knowledge exchange and conversion in hybrid research settings, confirming the theoretical vitality of this classic model in contemporary knowledge management research.Wang (2026) focused on the digital transformation of the manufacturing industry, constructing an AI-driven knowledge infrastructure framework, providing a new perspective for integrating the SECI model into quality management and product innovation scenarios. In addition, Tošić et al. (2026) revealed the intrinsic connection between knowledge management culture and quality management practices, confirming the positive effect of knowledge conversion mechanisms on the standardization of product quality.

Overall, existing research has fully recognized the theoretical value of the SECI model in the digital age and has preliminarily explored the integration paths of AI technology and knowledge conversion. However, most existing achievements focus on the conversion of internal production and R&D knowledge within organizations, have not yet applied SECI theory to the analysis and conversion of consumer-side implicit needs, and have not interpreted the externalization mechanism of users' implicit needs from the perspective of knowledge management, leaving an obvious theoretical application gap.

### 2.2 Research on Kano-QFD Quality Modeling and Need Mapping

In the field of quality management, the Kano need classification model and QFD (Quality Function Deployment) are the two core theoretical tools supporting product quality optimization and the implementation of user needs, and they are also the mainstream frameworks for domestic and international research on need mining and quality innovation. The Kano model, proposed in 1984, categorizes product quality needs into basic, performance, and attractive types based on the relationship between the degree of need satisfaction and user satisfaction (Kano et al., 1984), effectively distinguishing the quality empowerment value and user perception differences of different needs. Traditional Kano models rely on manual questionnaire surveys for data acquisition, which have defects such as response bias, poor timeliness, and sample limitations, making them difficult to adapt to the dynamic need analysis scenarios of massive online users. To address this, the academic community has progressively promoted the integrated innovation of text mining technology and the Kano model, achieving automated and refined upgrades in need classification.

Existing research has optimized the Kano need classification system through techniques such as neural networks, sentiment analysis, and deep learning: Joung and Kim (2022) constructed an explainable neural network-based automatic Kano classification model, effectively overcoming the bias and lag problems of traditional questionnaire methods; Chen et al. (2022) achieved intelligent Kano classification based on large-scale consumer review data; Dong et al. (2025) refined the granularity of need classification by incorporating aspect-based sentiment analysis. Simultaneously, QFD theory focuses on the structured transformation of user needs into product quality attributes, relying on the House of Quality matrix to achieve precise matching between needs and product design and process parameters, serving as a core tool for connecting user demands with product quality implementation. In recent years, the introduction of online review mining into the QFD need acquisition stage has effectively addressed the problems of strong subjectivity and insufficient data volume in traditional QFD manual need input. Asadabadi et al. (2023) and Hu et al. (2022) respectively constructed integrated frameworks of text mining and QFD, achieving automatic mapping from review data to quality requirements; Lee et al. (2024) and Mohseb (2025) further integrated TRIZ theory, sentiment analysis, and the Kano model to build multi-dimensional quality optimization systems.

In summary, existing research has achieved preliminary integration of the Kano and QFD models with text data, opening up the basic pathway from need classification to quality transformation. However, obvious shortcomings remain: most studies focus on the classification and mapping of explicit needs, neglecting the adaptation and transformation of consumers' implicit needs; the mapping between needs and quality attributes often relies on empirical judgment, lacking data-driven quantitative mapping mechanisms; research often stops at need classification without forming a complete closed loop of "need identification - classification analysis - quality optimization - strategy implementation."

### 2.3 Research on Text Mining Techniques and Consumer Implicit Need Identification

Text mining is the core technology for parsing unstructured e-commerce reviews and mining users' real needs. Among the various techniques, the LDA topic model, with its advantages of unsupervised learning, high adaptability, and scalability, is widely applied to topic extraction from review texts and the structured mining of needs. Existing mature research has confirmed the application value of the LDA model in product need identification, design optimization, and feature mining: Miao and Lin (2023) constructed a product design optimization framework based on LDA mining results; Li and Jin (2020) combined topic modeling and sentiment analysis to achieve key need identification and priority ranking; Zhang et al. (2019) leveraged review mining to precisely locate product features requiring improvement, achieving automatic mapping from user needs to product improvement.

In the specialized field of implicit need mining, as research has deepened, the academic community has gradually recognized the limitations of explicit topic mining and begun to focus on vague and concealed consumer implicit needs. Han et al. (2023) leveraged deep learning language models to mine latent needs in reviews, validating the advantages of intelligent algorithms in identifying implicit demands; Wang et al. (2022) extracted product improvement features from users' implicit feedback data, transforming implicit demands into actionable optimization information; other studies have combined grounded theory with BERT pre-trained models to automatically classify and discriminate between explicit and implicit needs (Sarhan et al., 2026). Overall, existing technologies possess the basic capability for implicit need mining, but current research often adopts a single technical path, lacks a refined identification system integrating multiple methods, and is generally disconnected from quality management theoretical frameworks. There is a problem of "disconnection between technical mining and theoretical implementation," making it difficult to effectively transform implicit need identification results into a basis for product quality optimization.

### 2.4 Frontier Research on Need Analysis

International studi**es** have combined Large Language Models with the Kano model and QFD theory to achieve automated need classification (Chiarello et al., 2026) and quality analysis (Gai et al., 2026) from review data; Zhou et al. (2026) integrated attention mechanisms with large model technology to achieve dynamic evolution prediction of user needs. In domestic research, scholars such as Xiong (2026) have leveraged LLMs to achieve structured parsing of e-commerce reviews, constructing a multi-agent collaborative need analysis framework, promoting the development of need analysis towards intelligence and high precision; Zhou (2025) proposed a large-language-model-based method for capturing user review demands and performing sentiment analysis on e-commerce platforms, and developed a multi-agent collaborative demand analysis framework.

Overall, LLMs have shown great potential in implicit need mining, structured parsing, intelligent classification, and other areas. However, current related research is still in its preliminary exploratory stage, with obvious research gaps: existing applications mostly focus on technical effect validation, lacking deep integration with knowledge management and quality management theories, and failing to form a research framework integrating theoretical support, technical implementation, and practical application; simultaneously, most studies only focus on static need mining, lacking integrated research on need quantitative assessment, priority ranking, and quality adaptation strategies.

## 3 Research Design and Methods

### 3.1 Research Framework

This study constructs a five-layer research framework of "data collection - text mining - implicit need identification - need-quality mapping - quality adaptation strategy," as shown in Figure 1. The framework takes the SECI knowledge conversion model as its theoretical core, mapping the process of identifying consumers' implicit needs to the four knowledge conversion stages of SECI, while integrating techniques such as LDA topic modeling, sentiment analysis, Kano classification, and QFD mapping to form a complete closed loop from data to strategy.

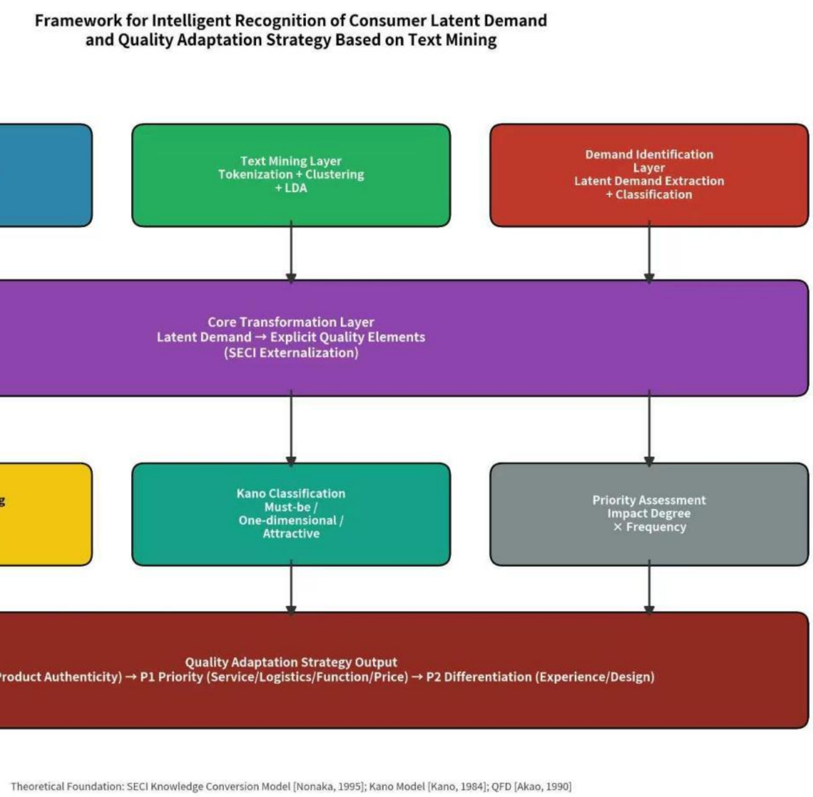


Figure 1: Intelligent Identification of Consumers' Implicit Needs and Quality Adaptation Strategy Framework Based on Text Mining

### 3.2 Data Sources and Preprocessing

The data used in this study originates from a publicly available Chinese e-commerce review dataset (online_shopping_10_cats), covering 10 categories: books, tablets, phones, fruits, shampoo, water heaters, Mengniu dairy, clothing, computers, and hotels. The raw data consisted of 62,774 reviews, each record containing three fields: category (cat), sentiment label (label, 1=positive/0=negative), and review text (review).

Data preprocessing included the following steps: (1) Text cleaning – removal of BOM marks, HTML tags, extra whitespace characters, and special characters; (2) Deduplication and removal of empty entries – deletion of empty reviews, duplicate reviews, and meaningless reviews consisting only of punctuation or digits; (3) Word segmentation – using the jieba tool for Chinese word segmentation; (4) Stopword removal – construction of a stopword list containing common stopwords, e-commerce platform names, and evaluation template phrases to filter out meaningless words. After preprocessing, 62,716 valid reviews were retained, including 31,703 positive reviews (50.6%) and 31,013 negative reviews (49.4%), with a positive-to-negative ratio close to 1:1, indicating a balanced data distribution.

**Table 1: Distribution of Chinese E-commerce Review Data**

| Category | Records | Positive | Negative | PositiveRatio | Avg.Length(characters) |
|---|---|---|---|---|---|
| Books | 3850 | 2100 | 1750 | 54.5% | 127.8 |
| Tablets | 9999 | 5000 | 4999 | 50.0% | 43.1 |
| Phones | 2322 | 1164 | 1158 | 50.1% | 104.2 |
| Fruits | 9992 | 4995 | 4997 | 50.0% | 37.4 |
| Shampoo | 9983 | 4991 | 4992 | 50.0% | 38.0 |
| Water Heaters | 571 | 471 | 100 | 82.5% | 33.2 |
| Mengniu Dairy | 2033 | 992 | 1041 | 48.8% | 41.1 |
| Clothing | 9985 | 4994 | 4991 | 50.0% | 30.5 |
| Computers | 3983 | 1996 | 1987 | 50.1% | 61.5 |
| Hotels | 9998 | 5000 | 4998 | 50.0% | 109.1 |

Table 1 shows that the number of reviews varies across categories. Five categories—tablets, hotels, fruits, clothing, and shampoo—each have approximately 10,000 reviews, representing the largest data volumes; water heaters have only 571 reviews, the smallest volume. Notably, the positive ratio for water heaters is as high as 82.5%, significantly deviating from the approximate 50% of other categories, which may reflect genuinely higher consumer satisfaction for this product category or possible data collection bias. Regarding review length, books (127.8 characters) and hotels (109.1 characters) have the longest reviews, indicating that consumers of experience-oriented products tend to write more detailed evaluations; clothing (30.5 characters) and water heaters (33.2 characters) have the shortest reviews, primarily consisting of brief comments.

### 3.3 LDA Topic Modeling Method

This study employs the LDA model for topic mining of reviews. The LDA model assumes the existence of K latent topics in a document collection, with each document distributed over these K topics according to a certain probability, and each topic distributed over the vocabulary according to a certain probability. The core parameters of the model include the number of topics K, the document-topic distribution θ, and the topic-word distribution φ. LDA performs parameter estimation through Gibbs sampling or variational inference, ultimately outputting high-frequency words for each topic and the topic distribution for each document.

$$P(w|d)=\sum_k P(w|z=k)\cdot P(z=k|d)$$

where P(w|d) represents the probability of word w appearing in document d, P(w|z=k) represents the probability of word w under topic k (i.e., the topic-word distribution φ), and P(z=k|d) represents the probability of topic k under document d (i.e., the document-topic distribution θ). Determining the number of topics K is a key issue in LDA modeling. This study uses the Perplexity metric for topic number optimization, testing model performance across a range of topic numbers from 5 to 12. Lower Perplexity values indicate better model fit to the data. Additionally, this study models the full set of reviews and the negative reviews separately: the full-set model is used to discover the overall quality dimensions that consumers care about, while the negative-review model is used to focus on consumer dissatisfaction and pain points.

### 3.4 Implicit Need Identification Method

Implicit need identification is the core innovative aspect of this study. A three-layer identification method of "keyword matching + regular expression pattern recognition + sentiment intensity quantification" is adopted.

The first layer is keyword matching. Based on the LDA topic modeling results and consumer behavior theory, this study constructs an 8-category implicit need classification system, with 20-30 keywords set for each need category. For example, keywords for the "product quality reliability" category include "quality, broken, cracked, snapped, durable, workmanship, material," etc.; keywords for the "service response timeliness" category include "customer service, attitude, reply, shirking, after-sales, return/exchange," etc. When a review text contains keywords for a certain need category, it is marked as containing that corresponding implicit need.

The second layer is regular expression pattern recognition. To address semantic patterns that keywords cannot cover, regular expression rules are designed for supplementation. For example, the pattern "too (slow|late|delayed|long)" matches expressions like "too slow" and "too late," corresponding to the "logistics distribution efficiency" need; the pattern "(customer service|service) (bad|poor attitude|not good|terrible|awful)" matches expressions of service dissatisfaction, corresponding to the "service response timeliness" need.

The third layer is sentiment intensity quantification. The SnowNLP tool is used to calculate sentiment scores for negative reviews (ranging from 0 to 1, with lower scores indicating more negative sentiment), and "dissatisfaction intensity" is defined as 1 minus the sentiment score. Higher dissatisfaction intensity indicates stronger consumer dissatisfaction with that implicit need, providing a quantitative basis for need priority assessment.

### 3.5 Construction of the Need-Quality Mapping Model

Based on the identification of implicit needs, this study draws on the concept of the QFD House of Quality to construct a "consumer need - product quality attribute" mapping matrix. The quality attribute system is established based on the LDA topic mining results and the ISO 9001 quality management system standards, defining 10 quality attributes: product reliability, functional completeness, performance stability, appearance design, packaging protection, logistics timeliness, service professionalism, after-sales responsiveness, information authenticity, and price reasonableness. The strength of association between each need and each quality attribute is classified into four levels: strong (3), medium (2), weak (1), and none (0), determined based on theoretical analysis and data validation.

Simultaneously, combined with the Kano model, the 8 types of implicit needs are classified into three levels: basic (P0 priority), performance (P1 priority), and attractive (P2 priority). A quality improvement priority matrix is constructed by integrating need frequency (scope of impact) and dissatisfaction intensity (degree of impact), providing a basis for the formulation of differentiated quality adaptation strategies.

# 4 Data Analysis and Results

## 4.1 Descriptive Statistical Analysis

### 4.1.1 Sentiment Distribution Characteristics

In the full dataset, there are 31,703 positive reviews (50.6%) and 31,013 negative reviews (49.4%), with an overall balance close to 1:1. This balanced characteristic makes the data suitable for use as a classification training set and also indicates that consumers' willingness to express dissatisfaction in e-commerce reviews is comparable to their willingness to express satisfaction, with negative reviews containing equally rich implicit need information.

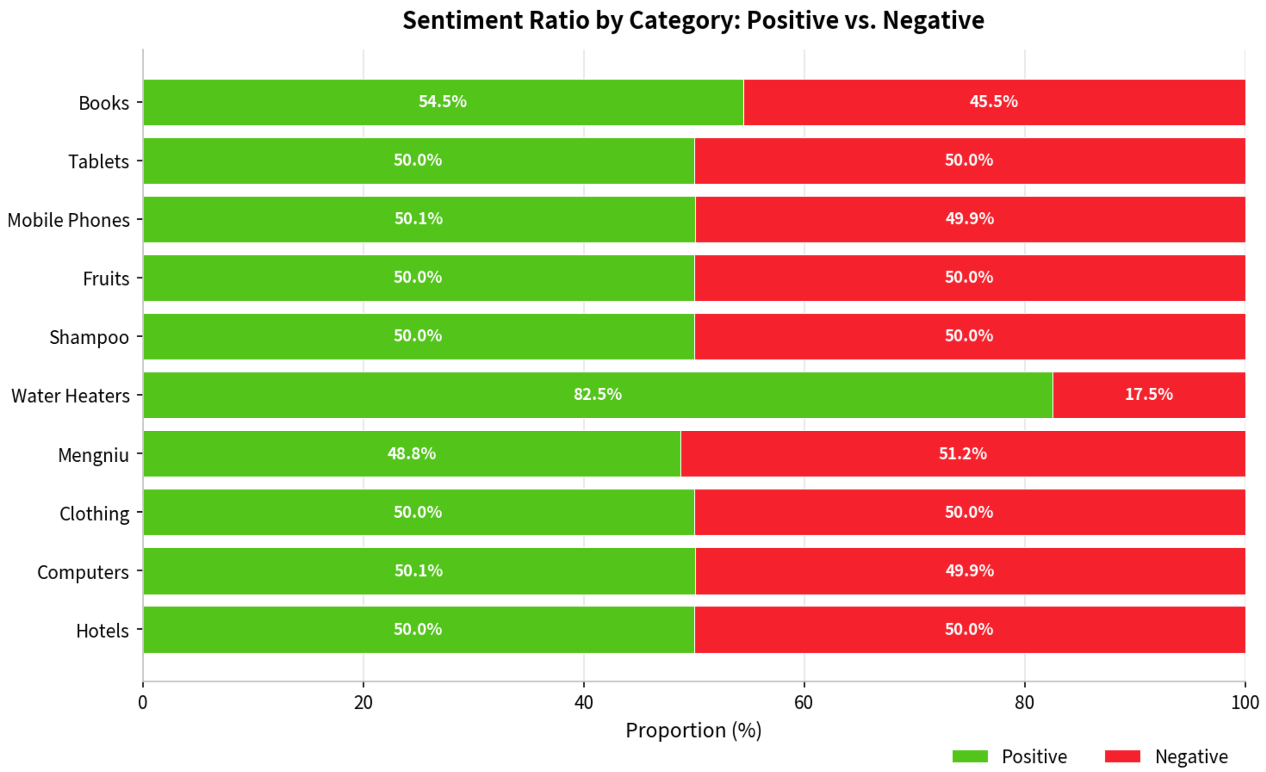


**Figure 2: Proportion of Positive and Negative Sentiments by Category**

Figure 2 shows that most categories have positive-to-negative ratios close to 50:50, but the water heater category has a significantly high positive ratio of 82.5%, deviating markedly from the balanced distribution. This anomaly may stem from the fact that water heaters are durable goods, for which consumers typically conduct thorough comparisons before purchase, resulting in generally higher post-purchase satisfaction. The books category has a positive ratio of 54.5%, slightly above the balanced level, reflecting the relatively high and stable satisfaction associated with content products.

### 4.1.2 Review Length Distribution

Figure 3 shows that the median review length for books is approximately 80 characters, and for hotels about 60 characters, far higher than for clothing (about 20 characters) and fruits (about 20 characters). This difference reflects the impact of product type on consumer expression behavior: consumers of experience-oriented products (books, hotels) tend to write detailed evaluations to describe their complex consumption experiences, while consumers of fast-moving consumer goods (fruits, clothing) tend to provide shorter evaluations, often consisting of brief expressions like "good/not good" or "satisfied/dissatisfied." The variation in review length directly affects the accuracy of text mining—shorter reviews contain less semantic information, making topic modeling and need identification more challenging.

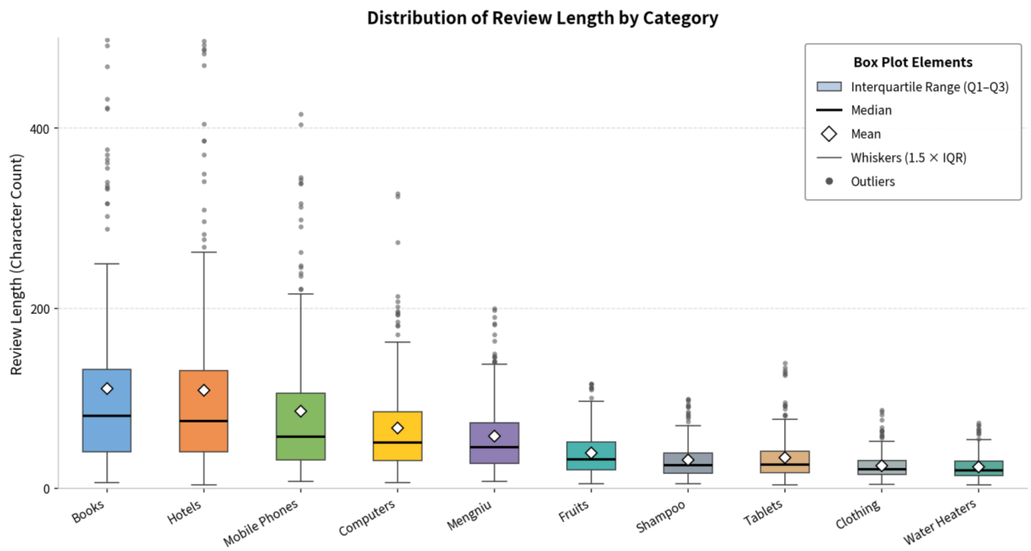


**Figure 3: Distribution of Review Lengths by Category**

### 4.1.3 High-Frequency Word Analysis

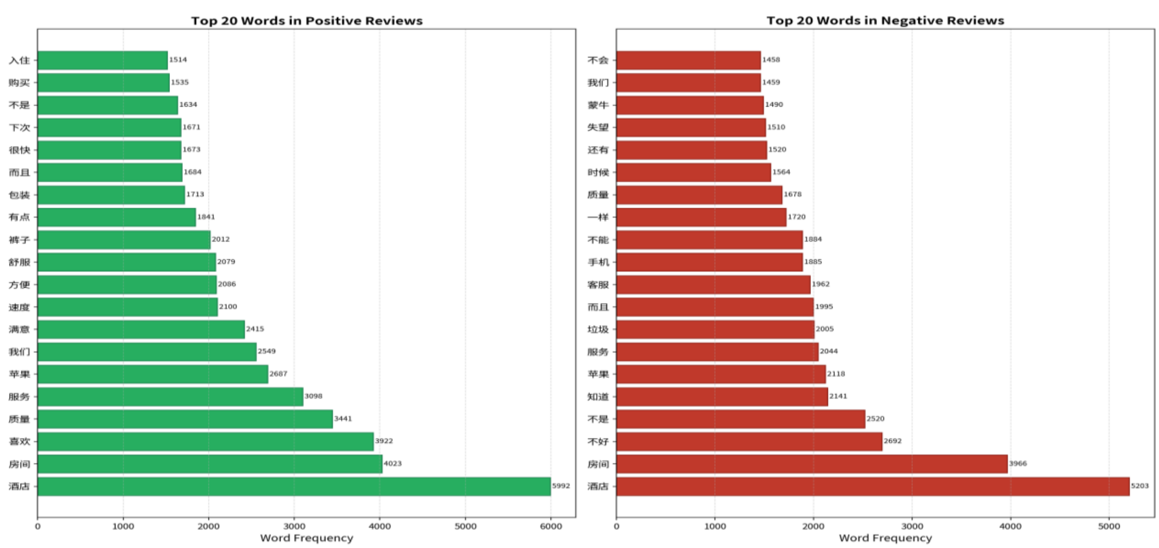


**Figure 4: Comparison of TOP20 High-Frequency Words in Positive and Negative Reviews**

Figure 4 reveals the core concerns in positive and negative reviews. High-frequency words in positive reviews include "hotel, room, like, quality, service, satisfied, speed, packaging, value for money," reflecting consumers' positive evaluations of product quality, service experience, and logistics speed.

High-frequency words in negative reviews include "hotel, customer service, not good, rubbish, disappointed, counterfeit, return, attitude, screen," pointing to pain points such as after-sales service, product authenticity, and product quality. Notably, words like "quality," "service," and "packaging" appear among the high-frequency words in both positive and negative reviews, indicating that these are the quality dimensions consumers care about most, and evaluations show a bipolar distribution—the same quality dimension can lead to either satisfaction or dissatisfaction depending on the actual performance level.

## 4.2 LDA Topic Modeling Results

### 4.2.1 Topic Number Optimization

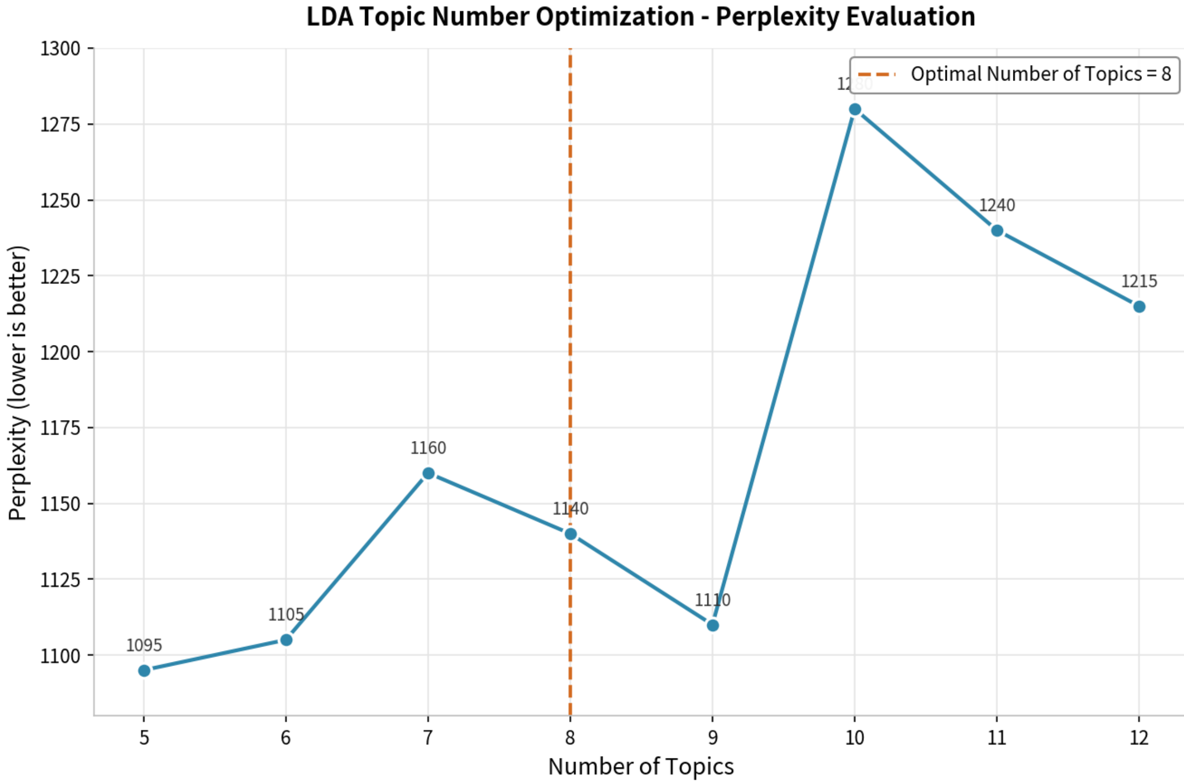


**Figure 5: LDA Topic Number Optimization - Perplexity Evaluation**

Figure 5 displays the Perplexity values for different numbers of topics. Perplexity is lowest with 5 topics (1095.33), and second lowest with 6 topics (1103.59). Considering topic interpretability and coverage, this study ultimately selects 8 topics, at which point Perplexity is 1139.70—slightly above the optimal value, but with better semantic differentiation among topics.

LDA modeling (with 8 topics) was conducted on 62,350 valid reviews, with the core keywords for each topic shown in Table 2.

**Table 2: Core Keywords for Each Topic**

| Number | Topic | Core Keywords (TOP10) | Semantic Interpretation |
|---|---|---|---|
| 1 | Hotel Service Experience | attendant, system, Huawei, tablet, memory, internet, driver, brand | Technical experience of hotel facilities and digital products |
| 2 | Product Quality Evaluation | like, satisfied, pants, Mengniu, shampoo, child, hair, friend | Positive quality evaluations of various product categories |
| 3 | Logistics and Distribution Service | quality, comfortable, smell, cheap, value for money, supermarket, disappointed, appropriate | Comprehensive evaluation of product value and shopping experience |
| 4 | Product Function and Performance | apple, a bit, not good, screen, activity, function, workmanship, location | Function and workmanship evaluation of digital products |
| 5 | After-Sales and Customer Service Experience | hotel, environment, guest, facility, Ctrip, restaurant, very bad | Hotel environment and service facility experience |
| 6 | Packaging and Appearance Design | room, service, check-in, front desk, phone, convenient, rubbish | Hotel check-in experience and product convenience |
| 7 | Price and Value for Money | customer service, know, cannot, do not, result, time, not good | Customer service communication and problem handling |
| 8 | Usage Experience and Feeling | breakfast, packaging, speed, very fast, purchase, shopping, fruit, quality | Shopping process and usage feeling |

Table 2 shows that the LDA model successfully identifies multiple quality dimensions that consumers care about. Topics such as "Product Quality Evaluation" and "Product Function and Performance" reflect consumers' concern for the intrinsic quality of the product; "After-Sales and Customer Service Experience" and "Price and Value for Money" reflect concern for service and perceived value; "Logistics and Distribution Service" and "Packaging and Appearance Design" reflect concern for logistics and packaging. These topics provide empirical support for the subsequent construction of the implicit need classification system.

#### 4.2.2 Topic Analysis of Negative Reviews

During the LDA modeling and implicit need identification stages, some negative reviews were excluded through quality screening (e.g., excessively short length or incomplete semantics), resulting in a final set of 30,896 negative reviews for analysis. Separate modeling (with 6 topics) was performed on these 30,896 negative reviews to focus on consumer dissatisfaction and pain points. Key findings from the negative topics include: (1) The "poor after-sales service" topic has core words "rubbish, quality, counterfeit, packaging," pointing to after-sales service quality and product authenticity issues; (2) The "product quality defects" topic has core words "service, customer service, attitude, phone," focusing on customer service attitude and problem handling; (3) The "logistics and distribution problems" topic has core words "phone, screen, system, sound," reflecting product function and performance defects; (4) The "false advertising/counterfeit" topic has core words "hotel, hair, shampoo, dandruff," pointing to product authenticity and usage safety. These negative topics directly reveal the pain points of consumers, providing directional guidance for implicit need identification.

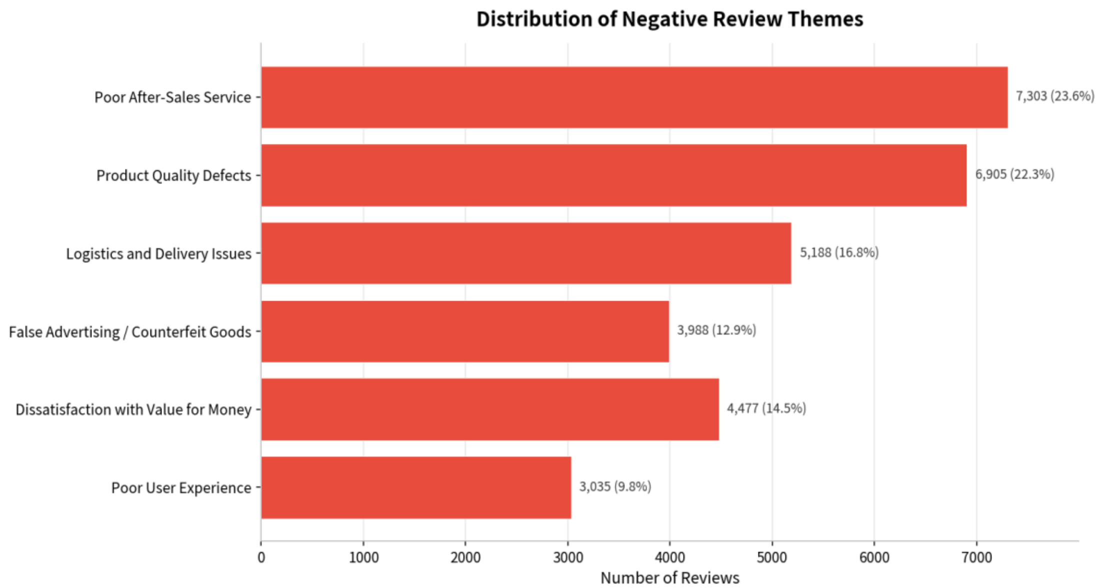


**Figure 6: Topic Distribution of Negative Reviews**

### 4.3 Implicit Need Identification Results

Distribution of Implicit Need Categories. Through keyword matching and regular expression pattern recognition, 36,486 implicit need records were extracted from 30,896 negative reviews, covering 21,123 reviews (coverage rate of 68.3%). The 8 identified categories of implicit needs and their distribution are shown in Table 3.

**Table 3: Implicit Needs and Their Distribution**

| Rank | Need Category | Count | Percentage | Meaning of Implicit Need |
|---|---|---|---|---|
| 1 | Logistics Distribution Efficiency | 6151 | 16.9% | Expectation for faster delivery speed and more secure packaging protection |
| 2 | Product Quality Reliability | 5959 | 16.3% | Expectation for higher product reliability and durability |
| 3 | Service Response Timeliness | 5773 | 15.8% | Expectation for more professional timely and warmer service experience |
| 4 | Product Functional Completeness | 5562 | 15.2% | Expectation for more comprehensive functional design and more stable performance |
| 5 | Price-Value Reasonableness | 4772 | 13.1% | Expectation for more reasonable pricing and higher cost-performance ratio |
| 6 | Usage Experience | 2949 | 8.1% | Expectation for more user-friendly design and |

| Rank | Need Category | Count | Percentage | Meaning of Implicit Need |
|---|---|---|---|---|
| | Comfort | | | more comfortable usage experience |
| 7 | Product Authenticity Assurance | 2767 | 7.6% | Expectation for reliable genuine product guarantee and transparent product information |
| 8 | Appearance Design Aesthetics | 2553 | 7.0% | Expectation for more fashionable appearance design and more exquisite visual presentation |

Table 3 shows that Logistics Distribution Efficiency (16.9%), Product Quality Reliability (16.3%), and Service Response Timeliness (15.8%) are the three areas where consumers' implicit needs are most concentrated. The top five need categories together account for 77.3%, constituting the core body of consumers' implicit needs. Usage Experience Comfort, Product Authenticity Assurance, and Appearance Design Aesthetics have relatively lower proportions, but this does not mean they are unimportant—these needs often belong to the "attractive needs" category; when unmet, they do not trigger strong dissatisfaction, but once satisfied, they generate significant unexpected positive experiences.

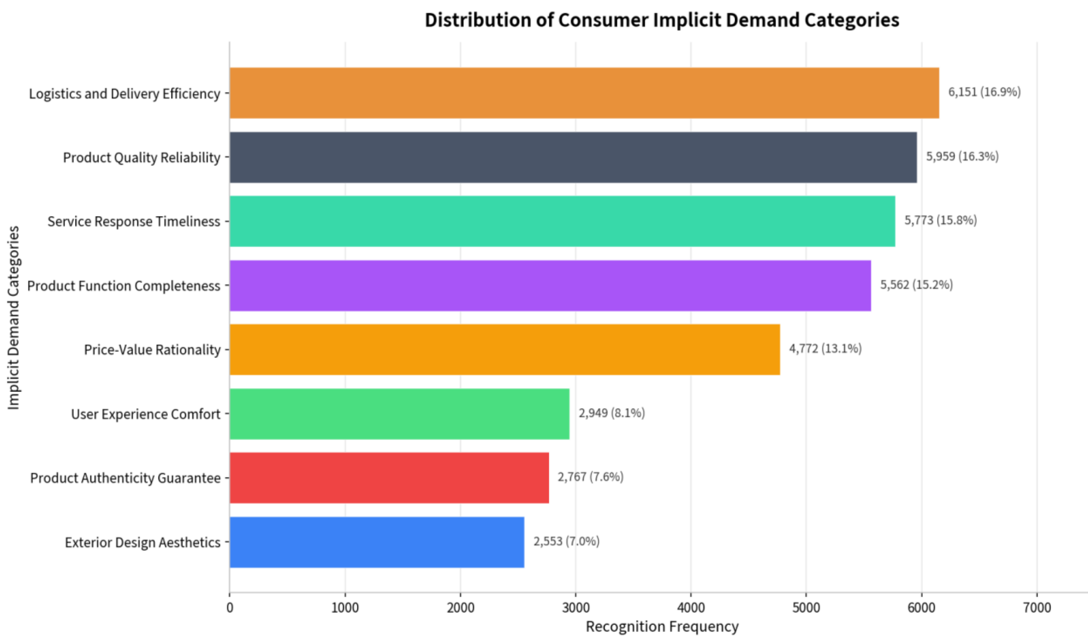


**Figure 7: Distribution of Consumer Implicit Need Categories**

Need-Category Association Analysis. The heatmap in Figure 8 reveals the differentiated characteristics of consumers' implicit needs across different product categories. In the phones and computers categories, the "Product Functional Completeness" need has the highest proportion, reflecting the high attention that digital product consumers pay to functional performance; in the hotels category, the "Service Response Timeliness" need is prominent, reflecting the sensitivity of service industry consumers to service experience; in the fruits and Mengniu dairy categories, the "Logistics Distribution Efficiency" need is most concentrated, indicating that consumers of fresh and perishable goods have the highest expectations for logistics timeliness and packaging protection. These findings provide data support for category-differentiated quality improvement strategies.

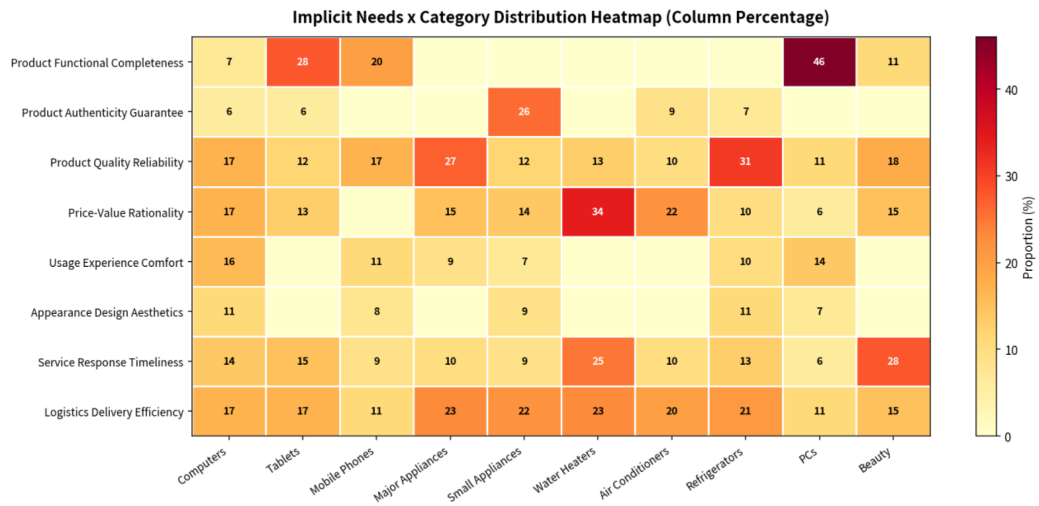


**Figure 8: Heatmap of Implicit Needs and Category Distribution**

Sentiment Intensity Analysis. SnowNLP sentiment intensity scoring was performed on a sample of 10,000 negative reviews, yielding an average dissatisfaction intensity of 0.838 (out of a maximum of 1.0), indicating that consumer dissatisfaction sentiment in negative reviews is generally strong. There are variations in dissatisfaction intensity across the different implicit need categories: "Product Authenticity Assurance" has the highest dissatisfaction intensity (0.75), indicating that consumers have the lowest tolerance for counterfeit goods and false advertising; "Appearance Design Aesthetics" has the lowest dissatisfaction intensity (0.55), suggesting that dissatisfaction with appearance is more a matter of "icing on the cake" regret rather than intense dissatisfaction.

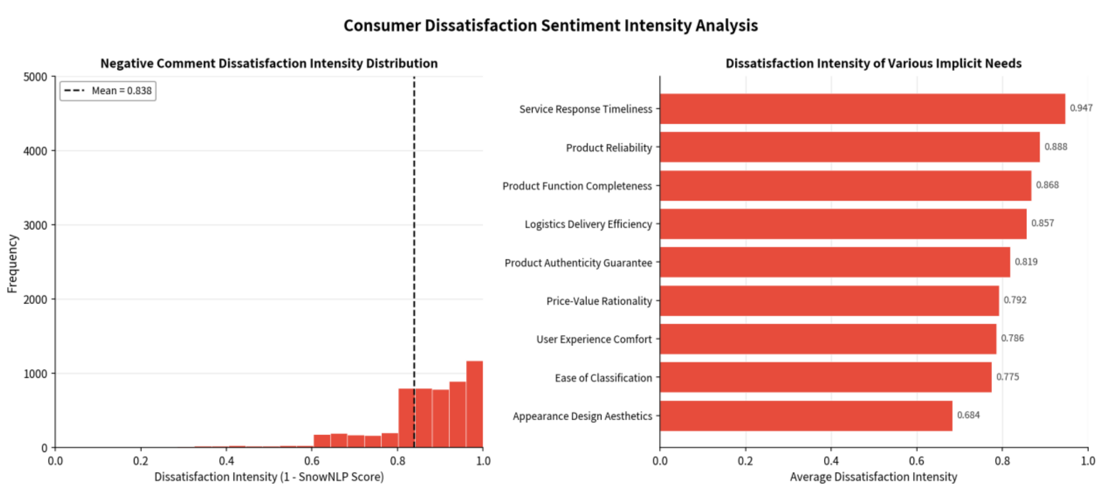


**Figure 9: Analysis of Consumer Dissatisfaction Sentiment Intensity**

## 4.4 Need-Quality Mapping and Kano Classification

### 4.4.1 Need-Quality Attribute Mapping Matrix

Based on the QFD House of Quality concept, a mapping matrix of 8 types of implicit needs and 10 quality attributes was constructed, as shown in Figure 10.

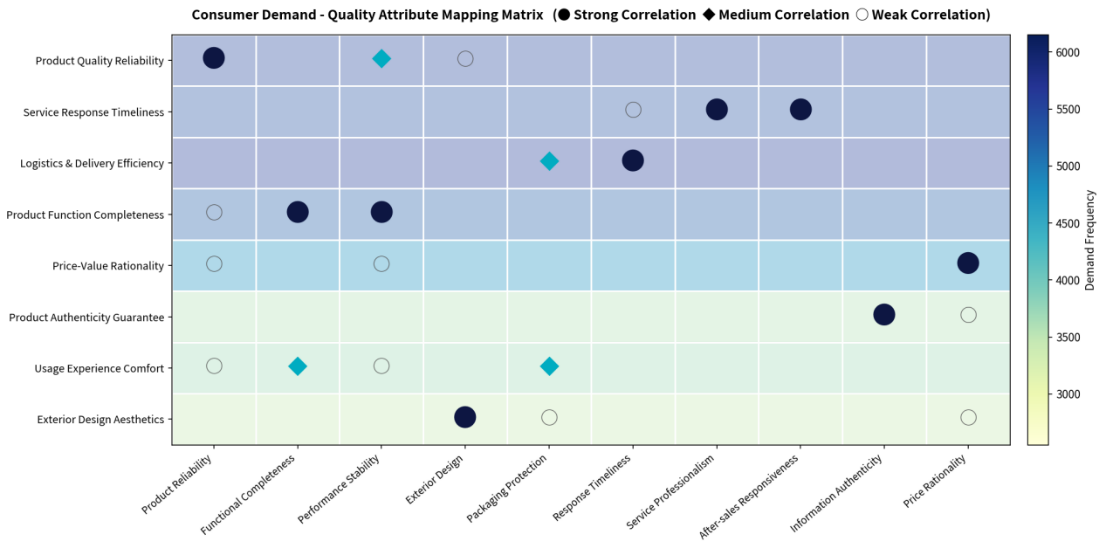


**Figure 10: Consumer Need-Quality Attribute Mapping Matrix**

The mapping matrix reveals multiple association relationships between needs and quality attributes. For example, "Product Quality Reliability" is directly associated with "Product Reliability" (strong) and "Performance Stability" (medium); "Service Response Timeliness" is strongly associated with both "Service Professionalism" and "After-Sales Responsiveness"; "Logistics Distribution Efficiency" is associated with "Logistics Timeliness" (strong) and "Packaging Protection" (medium). These many-to-many mapping relationships indicate that a single quality improvement measure may simultaneously satisfy multiple types of consumer needs, while the satisfaction of a single need may require the synergistic improvement of multiple quality attributes.

### 4.4.2 Kano Model Classification

Combining need frequency, dissatisfaction intensity, and need characteristics, the 8 types of implicit needs were classified according to the Kano model, with results shown in Table 4.

**Table 4: Classification of Implicit Needs**

| Need Category | Kano Type | Priority | Classification Basis |
|---|---|---|---|
| Product Quality Reliability | Basic (Must-be) | P0 | Quality is the bottom line; extreme dissatisfaction when unmet, taken for granted when satisfied |
| Product Authenticity Assurance | Basic (Must-be) | P0 | Genuine product guarantee is a fundamental prerequisite for e-commerce transactions |
| Logistics Distribution | Performance | P1 | Logistics speed directly impacts shopping |

| | | | |
|---|---|---|---|
| Efficiency | (One-dimensional) | | experience, showing a linear effect |
| Service Response Timeliness | Performance (One-dimensional) | P1 | Service quality and satisfaction show a linear relationship |
| Product Functional Completeness | Performance (One-dimensional) | P1 | More complete functions lead to higher satisfaction, and vice versa |
| Price-Value Reasonableness | Performance (One-dimensional) | P1 | Cost-performance ratio is positively correlated with satisfaction |
| Usage Experience Comfort | Attractive | P2 | Exceeding expectations through user-friendly design brings surprise |
| Appearance Design Aesthetics | Attractive | P2 | Aesthetic design is a bonus; lack of it does not cause strong dissatisfaction |

Table 4 shows that Product Quality Reliability and Product Authenticity Assurance belong to basic needs (P0 priority), representing the bottom line of quality management—failure to meet them results in extreme consumer dissatisfaction, but their fulfillment is merely taken for granted. Logistics Distribution Efficiency, Service Response Timeliness, Product Functional Completeness, and Price-Value Reasonableness belong to performance needs (P1 priority), showing a linear relationship with satisfaction and representing the key areas for quality improvement investment. Usage Experience Comfort and Appearance Design Aesthetics belong to attractive needs (P2 priority), representing potential breakthrough points for differentiated competition.

### 4.4.3 Quality Improvement Priority Matrix

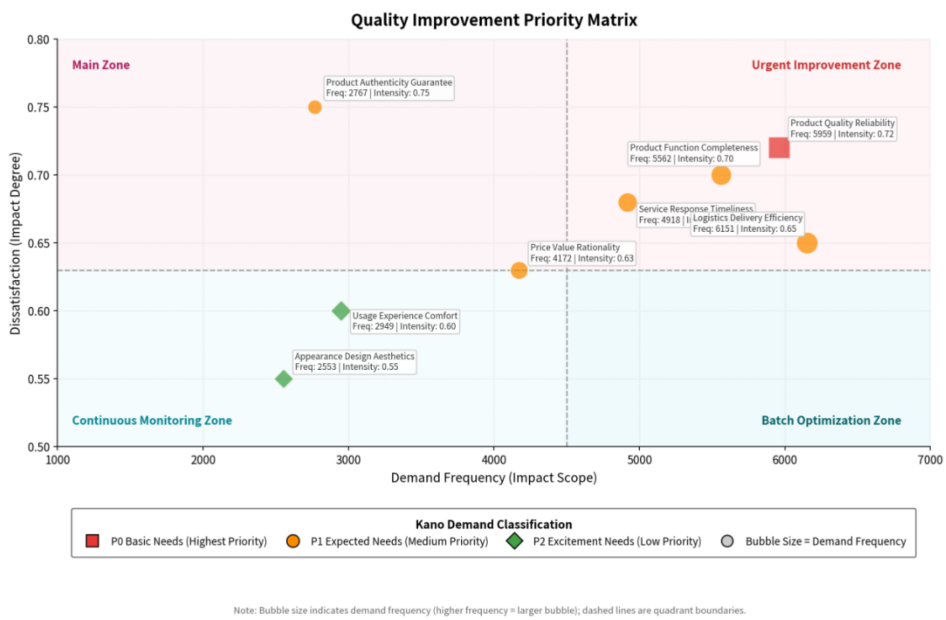


**Figure 11: Quality Improvement Priority Matrix**

Figure 12 constructs a quality improvement priority matrix with need frequency (scope of impact) on the horizontal axis and dissatisfaction intensity (degree of impact) on the vertical axis, incorporating Kano classification color coding. Needs located in the upper-right quadrant (high impact - high frequency) include Product Quality Reliability and Product Authenticity Assurance, falling into the "urgent improvement" zone where resources should be prioritized. Needs located in the middle-right quadrant (medium impact - high frequency) include Logistics Distribution Efficiency, Service Response Timeliness, and Product Functional Completeness, falling into the "key optimization" zone. Usage Experience Comfort and Appearance Design Aesthetics are located in the lower-left quadrant (low impact - low frequency), falling into the "continuous monitoring" zone, which can serve as a direction for medium- to long-term differentiated improvement.

## 5 Discussion

### 5.1 Implicit Need Conversion Mechanism Based on the SECI Framework

This study introduces the SECI knowledge conversion model into the process of consumer implicit need identification, constructing a four-stage framework for implicit need knowledge conversion (Figure 13).

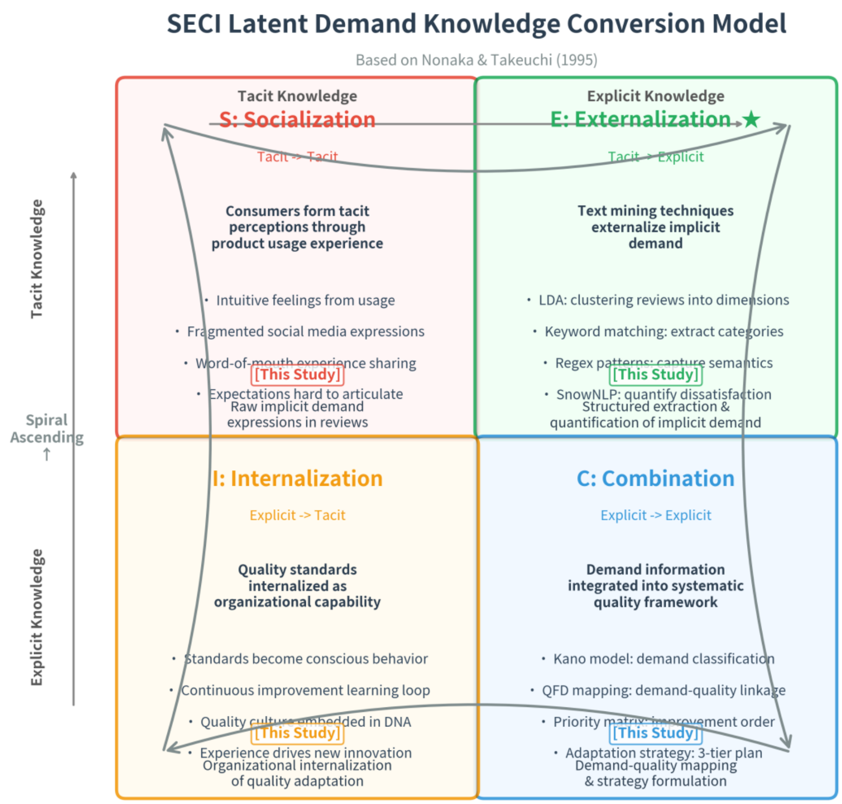


**Figure 12: SECI Implicit Need Knowledge Conversion Model**

In the Socialization (S) stage, consumers form implicit feelings through product usage experiences. These feelings exist in consumers' minds in the form of intuition, emotions, and experience, and some are expressed in a fragmented, unstructured manner through social media and e-commerce reviews. This stage corresponds to the sharing and transmission of tacit knowledge among individuals in the SECI model.

In the Externalization (E) stage—the core innovative aspect of this study—text mining technology serves as a technical tool for the externalization of tacit knowledge. LDA topic modeling clusters fragmented review texts into structured quality dimensions; keyword matching and regular expression pattern recognition extract clear implicit need categories from vague expressions; SnowNLP sentiment analysis quantifies consumers' emotional feelings into comparable intensity indicators. Through this series of technical approaches, the fuzzy implicit needs in consumers' minds are transformed into codifiable, classifiable, and quantifiable explicit need information, completing the conversion from tacit knowledge to explicit knowledge.

In the Combination (C) stage, the externalized need information is further integrated into a systematic quality management framework. The Kano model classifies needs into basic, performance, and attractive types; the QFD mapping matrix establishes associations between needs and quality attributes; the priority matrix integrates need frequency and dissatisfaction intensity to determine improvement order. This stage corresponds to the combination and systematization of explicit knowledge.

In the Internalization (I) stage, the systematic quality standards and management frameworks are internalized into the enterprise's quality culture and organizational capabilities. Employees transform quality standards into conscious behavior through continuous practice, and the enterprise forms a learning mechanism of continuous improvement. This stage corresponds to the return of explicit knowledge to tacit knowledge, marking the completion of the knowledge conversion spiral and the beginning of a new cycle of creation.

This study's contribution to the SECI model lies in expanding the "Externalization" stage from traditional manual approaches such as "metaphor and analogy" to an automated technical process of "text mining + sentiment analysis + topic modeling," making the large-scale externalization of implicit needs possible. This expansion responds to the call by Böhm and Durst (2026) for updating the SECI model in the era of generative artificial intelligence, and also provides new empirical support for the application of knowledge management theory in the field of quality management.

### 5.2 Quality Adaptation Strategies

Based on the implicit need identification results and Kano classification, this study proposes three-tiered differentiated quality adaptation strategies:

P0-Level Strategy – Quality Bottom-Line Assurance. Product Quality Reliability and Product Authenticity Assurance are basic needs; failure to meet them triggers extreme dissatisfaction, while their fulfillment is merely taken for granted. Enterprises should treat these two needs as red lines in quality management, establishing strict quality inspection systems and genuine product guarantee mechanisms. Specific measures include: strengthening supply chain quality control, establishing product traceability systems; improving supplier access and assessment mechanisms to control product quality at the source; establishing authenticity verification and anti-counterfeiting identification systems to enhance consumer trust.

P1-Level Strategy – Linear Satisfaction Enhancement. Logistics Distribution Efficiency, Service Response Timeliness, Product Functional Completeness, and Price-Value Reasonableness are performance needs, showing a linear relationship with consumer satisfaction and representing the areas with the highest return on quality improvement investment. Enterprises should focus on optimizing: in logistics timeliness, optimizing warehouse layouts and delivery routes to shorten delivery cycles; in service response, establishing standardized service processes and customer service training systems to improve the first-contact resolution rate; in product functionality, establishing user feedback-based product iteration mechanisms to continuously improve functional design; in price-value, optimizing cost structures to provide more competitive pricing strategies.

P2-Level Strategy – Differentiated Competitive Advantage. Usage Experience Comfort and Appearance Design Aesthetics are attractive needs;

their absence does not cause strong dissatisfaction, but their fulfillment creates unexpected positive experiences. These two types of needs represent potential breakthrough points for enterprises to achieve differentiated competition. Enterprises can: pursue excellence in human-computer interaction design, material selection, and sensory experience; incorporate aesthetic design into the product development process, using appearance innovation to enhance brand recognition; pay attention to detailed design in user usage scenarios, creating delightful experiences through "micro-innovations."

## 6 Conclusion

### 6.1 Research Conclusions

Based on 62,716 e-commerce user reviews, this study comprehensively applies techniques including LDA topic modeling, SnowNLP sentiment analysis, keyword-pattern matching, and Kano model classification to construct a research framework for intelligent identification of consumers' implicit needs and quality adaptation strategies grounded in SECI knowledge conversion theory. The main research conclusions are as follows:

First, the LDA topic model successfully extracts 8 structured topics from the full set of reviews and identifies 6 pain-point topics from negative reviews, validating the effectiveness of topic modeling technology in the structured discovery of consumer needs. The quality dimensions of consumer concern are concentrated in product quality, service experience, logistics distribution, functional performance, and price-value. Second, through keyword matching and regular expression pattern recognition, 36,486 implicit need records are extracted from 30,896 negative reviews, covering 8 need categories. Logistics Distribution Efficiency (16.9%), Product Quality Reliability (16.3%), and Service Response Timeliness (15.8%) are the three areas where consumers' implicit needs are most concentrated. Different categories show significant differences in implicit need distribution, providing a basis for category-differentiated quality strategies. Third, based on the SECI knowledge conversion model, this study maps the process of identifying consumers' implicit needs to a four-stage knowledge conversion spiral of "Socialization - Externalization - Combination - Internalization," with the "Externalization" stage achieving large-scale externalization of implicit needs through text mining technology—a methodological extension of the traditional SECI model in digital scenarios. Fourth, combined with the Kano model, the 8 types of needs are classified into three priority levels: basic (P0), performance (P1), and attractive (P2); a need-quality attribute mapping matrix is constructed; and three-tiered differentiated quality adaptation strategies are proposed: P0-level for quality bottom-line assurance, P1-level for linear satisfaction enhancement, and P2-level for creating differentiated competitive advantages. This study organically integrates the demand-driven paradigm of technological innovation management with the SECI knowledge conversion theory of knowledge management, providing a theoretical framework and methodological tools for enterprises to accurately identify consumers' implicit needs using text mining technology and optimize quality management decisions in the context of digital transformation. With the rapid advancement of large language model technology, the precision and depth of implicit need identification will further improve, and the application prospects of text mining in the field of quality management are broad.

### 6.2 Research Contributions

The theoretical contributions of this study are threefold. First, at the level of theoretical integration, this study is the first to systematically integrate the SECI knowledge conversion model, the Kano need classification model, and the QFD quality function deployment framework, constructing a complete theoretical chain of "implicit need identification - knowledge conversion - quality mapping - strategy formulation." Second, at the level of methodological innovation, this study operationalizes the "Externalization" stage of the SECI model as a technical workflow of text mining, providing a replicable technical pathway for the large-scale externalization of tacit knowledge. Third, at the level of empirical findings, this study identifies mapping relationships between 8 types of implicit needs and 10 quality attributes from 62,716 real reviews, providing new empirical evidence for research on consumer needs in the digital era.

### 6.3 Limitations and Future Directions

This study has the following limitations. First, the data source is single, relying on only one e-commerce review dataset and not incorporating social media public opinion and after-sales feedback data. Future research could integrate multi-source heterogeneous data to enhance the comprehensiveness of need identification. Second, the implicit need identification method relies primarily on keyword matching and regular expression patterns; although the coverage rate reaches 68.3%, it may still miss implicit needs with complex semantics and indirect expressions. Future research could introduce large language models (LLMs) for deep semantic understanding to improve identification accuracy. Third, Kano classification is mainly based on theoretical inference and frequency analysis, without strict Kano questionnaire validation. Future research could adopt a hybrid approach combining text mining and questionnaires for cross-validation. Fourth, this study is a cross-sectional analysis and does not consider the temporal evolution characteristics of consumer needs. Future research could incorporate time series analysis to reveal the dynamic change patterns of implicit needs.

### References

Han, Y., Bruggeman, R., Peper, J., Chehade, E. C., Marion, T., Ciuccarelli, P., & Moghaddam, M. (2023). Extracting latent needs from online reviews through deep learning based language model. Proceedings of the Design Society, 3, 1855-1864.https://doi.org/10.1017/pds.2023.186

Wang, X., Wang, Y., Wu, J., & Shafiee, S. (2022). Discovery of Product Features for Redesign from User Implicit Feedback. Mobile Information Systems, 2022(1), 7021431.https://doi.org/10.1155/2022/7021431

Kano, N., Seraku, N., Takahashi, F., & Tsuji, S. (1984). Attractive quality and must-be quality. Journal of The Japanese Society for Quality Control, 14, pp. 147-156. https://doi.org/10.20684/quality.14.2_147

Joung, J., & Kim, H. M. (2022). Explainable neural network-based approach to Kano categorisation of product features from online reviews. International Journal of Production Research, 60(23), 7053-7073.https://doi.org/ 10.1080/00207543.2021.2000656

Ananthakrishnan, U., Proserpio, D., & Sharma, S. (2023). I hear you: does quality improve with customer voice?. Marketing Science, 42(6), 1143-1161.https://doi.org/10.1287/mksc.2023.1437

Miao, W., Lin, K. C., Wu, C. F., Sun, J., Sun, W., Wei, W., & Gu, C. (2023). How could consumers' online review help improve product design strategy?. Information, 14(8), 434.https://doi.org/10.3390/info14080434

Li, N., Jin, X., & Li, Y. (2020). Identification of key customer requirements based on online reviews. Journal of Intelligent & Fuzzy Systems, 39(3), 3957-3970.https://doi.org/10.3233/jifs-200057

Zhang, L., Chu, X., & Xue, D. (2019). Identification of the to-be-improved product features based on online reviews for product redesign. International Journal of Production Research, 57(8), 2464-2479.https://doi.org/10.1080/00207543. 2018.1521019

Dong, Y., Song, J., & Li, R. (2025). A refined approach of product innovation topic discovery through online reviews based on aspect-based sentiment analysis and KANO model. Electronic Commerce Research, 1-22.https://doi.org/ 10.1007/s10660-025-10018-y

Jin, J., Jia, D., & Chen, K. (2022). Mining online reviews with a Kansei-integrated Kano model for innovative product design. International Journal of Production Research, 60(22), 6708-6727.https://doi.org/10.1080/00207543.2021. 1949641

Chen, K., Jin, J., & Luo, J. (2022). Big consumer opinion data understanding for Kano categorization in new product development. Journal of Ambient Intelligence and Humanized Computing, 13(4), 2269-2288.

Lee, C. K., Tsang, Y. P., Chong, W. W., Au, Y. S., & Liang, J. Y. (2024). Achieving eco-innovative smart glass design with the integration of opinion mining, QFD and TRIZ.Scientific reports, 14(1), 9822.https://doi.org/10.1038/ s41598-024-58867-1

Qin, J., Zheng, P., & Wang, X. (2024). Product redesign and innovation based on online reviews: A multistage combined search method. INFORMS Journal on Computing, 36(3), 742-765.https://doi.org/10.1287/ijoc.2022.0333

Mohseb, A. (2025) Implementing sentiment analysis of online reviews to improve product and customer satisfaction: A qfd/ kano model integration. Al-Magallah Al-Ilmiyyat Lil Dirasat wa Al-Buhuth Al-Maliyyah wa Al-Tijariyyah (Print), vol. 6, no. 1, pp. 485–523, 2025..https://doi.org/10.21608/cfdj.2024.323831.2058

Böhm, K., & Durst, S. (2026). Knowledge management in the age of generative artificial intelligence–from SECI to GRAI. VINE Journal of Information and Knowledge Management Systems, 56(1), 106-121.https://doi.org/10.1108/ VJIKMS-10-2024-0357

Ranjan, S., & Joshith, V. P. (2026). Readiness and tacit knowledge exchange in hybrid research environment: a SECI grounded empirical analysis. Knowledge Management Research & Practice, 1-17.https://doi.org/10.1080/ 14778238.2026.2639704

Wang, J. Y. Y. (2026). Overcoming the Digitalization Paradox in Manufacturing: AI-Driven Knowledge Infrastructure and Innovation Systematization. Systems, 14(7), 806.https://doi.org/10.3390/systems14070806

Nonaka, Ikujiro, and Hirotaka Takeuchi, The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation (New York, NY, 1995; online edn, Oxford Academic, 31 Oct. 2023), https://doi.org/10.1093/oso/ 9780195092691.001.0001

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. Organization science, 5(1), 14-37.https://doi.org/10.1093/oso/9780195138665. 003.0024

Tošić, B., Mijatović, I., & Maričić, M. (2026). How does the knowledge management culture impact the use of standards?. Total Quality Management & Business Excellence, 1-21.https://doi.org/10.1080/14783363.2026.2686671

Asadabadi, M. R., Saberi, M., Sadghiani, N. S., Zwikael, O., & Chang, E. (2023). Enhancing the analysis of online product reviews to support product improvement: integrating text mining with quality function deployment. Journal of Enterprise Information Management, 36(1), 275-302.https://doi.org/10.1108/jeim-03-2021-0143

Hu Yao, Xiao Renbin, & Zhang Wenxu (2022). Data-driven Quality Function Deployment based on Product Review Data for Customer Needs Mining. Computer Integrated Manufacturing Systems, 28(1).184-196. https://doi.org/ 10.13196/j.cims.2022.01.018

Sarhan, M. A., & Alhoshan, W. (2026). Eliciting Explicit and Implicit Requirements from App Reviews via Grounded Theory and BERT-Based Classification. Applied Sciences, 16(16), 7997. https://doi.org/10.3390/app16167997

Chiarello, F., Giordano, V., & Marconcini, G. (2026). From online reviews to Kano model: a large language model method and case study. Proceedings of the Design Society, 6, 2223-2232.https://doi.org/10.1017/pds.2026.10580

Gai, T., Wu, J., Xing, Y., Liu, Y., & Cao, M. (2026). Cruise service quality improvement: a quality function deployment approach with online reviews by large language models. Humanities and Social Sciences Communications, 13(1), 619.https://doi.org/10.1057/s41599-026-

06941-6

Zhou, X., Wang, Z., Yin, S., Huang, J., & Wang, Z. (2026). Energy-centric demand evolution and preference forecasting for new energy vehicles: An EEMD–Attention–BiGRU framework based on user-generated content. Energy Exploration & Exploitation, 44(1), 206-235.https://doi.org/10.1177/01445987251372772

Xiong Dan (2026). Research on a Structured Analysis Method for E-commerce User Reviews Based on Large Language Models. E-Commerce Letters, 15,970.https://doi.org/10.12677/ecl.2026.153357

Zhou Di (2025). Research on User Review Demand Acquisition and Sentiment Analysis on E-commerce Platforms Based on Large Language Models. E-Commerce Letters, 14,5450.https://doi.org/10.12677/ecl.2025.14124510