Information Research - Vol. 31 No. iConf (2026)

Understanding user experience in generative AI application: Evidence from topic modelling and sentiment analysis of user reviews

Chenxin Zhou and Lihong Zhou

DOI: *https://doi.org/10.47989/ir31iConf64160*

Abstract

Introduction. With the rapid surge of users in generative AI (GenAI) applications, user experience research is moving beyond surveys and experiments toward large-scale analyses of online reviews. Integrating topic and sentiment patterns from these reviews with user experience theory enables a clearer view of users' concerns and experience gaps.

Method. This study analysed Chinese user reviews of DeepSeek using BERTopic for topic modelling combined with sentiment analysis and mapped the findings onto Jesse James Garrett's five elements of user experience to reveal layered perceptions.

Results. Thirty-four topics were identified and grouped into nine categories covering technical stability, device adaptation, AI interaction and functional features. Negative sentiments centred on server instability, system errors and functional deficiencies, whereas positive sentiments highlighted AI performance, emotional support, and distinctive functions. The five-element mapping revealed strategic instability, scope plane functional gaps, structural interaction barriers and framework plane issues with multi-device adaptation.

Conclusion. Joint topic–sentiment analysis not only uncovers core concerns of GenAI application users but also offers actionable insights for improving technical stability, device adaptation and interaction design, providing a new empirical view for optimising similar applications and advancing user experience research.

Introduction

Generative artificial intelligence (GenAI) applications have rapidly emerged as a transformative force in the technological landscape, profoundly reshaping how people search for, create, and interact with information. For example, ChatGPT quickly amassed millions of downloads after its launch, triggering an explosive surge in GenAI mobile applications worldwide. In China, the domestically developed GenAI application DeepSeek has drawn significant attention due to its rapid growth and extensive user adoption, leading to a substantial volume of Chinese-language user reviews. This makes it an ideal case for exploring the user experience of GenAI applications within the Chinese context, offering authentic insights into user feedback that complement existing studies, which have primarily focused on English-language reviews.

User experience (UX) is a critical factor in mobile application development, influencing user acceptance, usability, and overall satisfaction (Lu et al., 2025). It encompasses emotional, sensory, and contextual dimensions that significantly impact user behaviour and retention (Alshammare et al., 2025).

Many scholars have investigated the user experience of generative AI applications through methods such as surveys, usability testing, and system evaluations, highlighting the significant influence of perceived usefulness, trust, and design appeal on users' adoption intentions (Alabduljabbar, 2024; Alhejji et al., 2022; Bubaš, et al., 2024). However, research focusing on users' actual usage scenarios and genuine feedback remains relatively underdeveloped. Recently, some studies have begun analysing user reviews from app stores to examine usage issues and preferences, demonstrating that such reviews can provide more direct and reliable insights into user experience (Hadwan et al., 2022). These findings indicate that leveraging user reviews represents a valuable approach for understanding user needs and informing the optimisation of generative AI applications. Yet, studies combining topic modelling with sentiment analysis of GenAI user reviews remain scarce, especially for non-English contexts.

In summary, existing research primarily focuses on model performance rather than genuine user experience, with most studies analysing feedback from English-speaking users whilst insights from Chinese users remain unexplored. As a representative Chinese generative AI application, DeepSeek provides an ideal case study for examining user experience within such contexts. Against this backdrop, this study takes DeepSeek as a case to examine how Chinese users perceive and evaluate GenAI applications. We first apply BERTopic to extract and cluster latent topics from unstructured review texts; then conduct sentiment analysis to measure the emotional valence of each topic; and finally interpret the findings systematically through the view of Garrett's Five Elements of User Experience—strategy, scope, structure, framework, and surface (Garrett, 2010), examining aspects such as technical stability, functional completeness, and interaction fluency. This mixed method approach bridges computational text analytics with a well-established UX framework, moving beyond fragmented feedback toward a structured diagnosis.

Accordingly, we propose the following research questions:

RQ1: What latent topics emerge from DeepSeek's Chinese user reviews?

RQ2: How do sentiment patterns across these topics reflect user satisfaction and dissatisfaction?

RQ3: How can the insights derived from topic and sentiment analysis be applied to inform practical UX enhancements?

By answering these questions, the study illustrates a replicable approach to extracting and interpreting user concerns from a dataset of Chinese-language app-store reviews, offering concrete guidance for improving the user experience of DeepSeek and similar GenAI applications.

Literature review

User experience (UX) is a key dimension for evaluating the success of information systems and products, encompassing users' perceptions, emotional responses, value cognition, and overall satisfaction throughout the usage process (Law et al., 2009). In generative AI applications, UX represents the overall quality of users' experiences during interaction with a product, including perceptual, emotional, and functional aspects. (Alabduljabbar, 2024; Gu et al., 2024) Understanding these multifaceted aspects is therefore essential for designing and optimising generative AI applications.

Among the various conceptual approaches, Jesse James Garrett's five elements of user experience provides a widely cited theoretical framework for structuring product design. This model divides the UX design process into five planes—strategy, scope, structure, framework, and surface—emphasising that an excellent UX must be built plane by plane from the bottom up and that omissions at any plane may affect the final user perception (Garrett, 2010). This model provides macroscopic design guidance for user experience research.

As generative AI applications become increasingly prevalent, research into their user experience has been growing. Existing studies have primarily focused on design principles and evaluation methods. They often use expert evaluations, literature reviews, and multiple rounds of heuristic assessments to establish design standards. For example, Weisz et al. (2024) proposed six design principles for generative AI applications based on literature reviews and expert feedback. Amershi et al. (2019) validated 18 guidelines in human-AI interaction through practitioner testing. However, such approaches are generally theory-driven and focused on expert perspectives, lacking sufficient attention to the dynamic changes in real user experience.

Against the backdrop of increasing academic focus on user generated content (UGC), user review analysis has become an essential methodology within user experience (UX) research. As a form of UGC, user reviews offer high authenticity, timeliness, and scalability, capturing users' real attitudes and emotions in natural usage scenarios (Wang & Liu, 2023). Compared with more conventional approaches like surveys and laboratory studies, user reviews supply large quantities of unfiltered, real-time feedback, which often enhances the external validity.

In recent years, natural language processing (NLP) techniques have gained prominence in UX research, especially those that integrate topic modelling techniques such as BERTopic and sentiment analysis. These techniques enable researchers to identify latent themes within large collections of unstructured user reviews and to assess the emotional responses associated with those themes. (Devlin et al., 2019). For example, BERTopic automatically identifies core issues and concerns discussed in reviews, while sentiment analysis helps understand users' emotional attitudes towards these issues. Studies show that combining these techniques effectively identifies usability issues, privacy concerns, update problems, and themes reflecting user satisfaction or dissatisfaction (Ahmed et al., 2022; Baj-Rogowska & Sikorski, 2023; Ossai & Wickramasinghe, 2023). These methods fit the aims of this study, especially for analysing Chinese generative AI products.

Compared with theory-driven research, empirical research based on user reviews provides more practical and authentic feedback. Shao et al. (2025) examined open-source generative AI mobile applications and reported widespread integration challenges that influenced key user experience factors, particularly functionality and security. Nahar et al. (2024) investigated how latency and energy consumption affect the integration of generative AI, using interviews and survey data. Chen et al. (2025) analysed posts from the OpenAI Developer Forum and identified challenges related to prompting, API usage, and plugin development in generative AI applications.

Academic studies have increasingly focused on user experiences with generative AI in various domains. Surveys by Golding et al. (2024) and Kim et al. (2025) explored its use and perception in academic contexts, revealing variations in experience across roles, gender, and disciplines. Similarly, Shata and Hartley (2025) applied the Technology Acceptance Model to assess teachers' perceptions and their effect on engagement.

While existing studies have provided valuable insights into the UX of generative AI applications, most research has focused on English-language reviews, and studies analysing Chinese user feedback remain limited. This study analyses Chinese-language user reviews using BERTopic topic modelling and sentiment analysis to identify latent topics and emotional tendencies, providing a comprehensive understanding of users' needs and pain points. The findings offer actionable recommendations for optimising generative AI applications.

Research methodology and process

This chapter outlines the methodology and process adopted in this study, providing a detailed account of the research design, data collection, and analytical procedures.

Research design

This study adopts a multi-stage methodology combining topic modelling, sentiment analysis, and theoretical mapping. Using Chinese user reviews from the DeepSeek App, it aims to uncover latent themes and sentiment patterns and interpret them through Garrett's five-element UX framework to generate actionable design insights. The methodological framework of this study consists of four stages: (1) collection and preprocessing of user reviews, (2) identification of themes through topic modelling, (3) evaluation of user attitudes by sentiment analysis, and (4) integration of findings with the five elements of user experience to guide product optimisation. The technical roadmap of this study is shown in Figure 1.

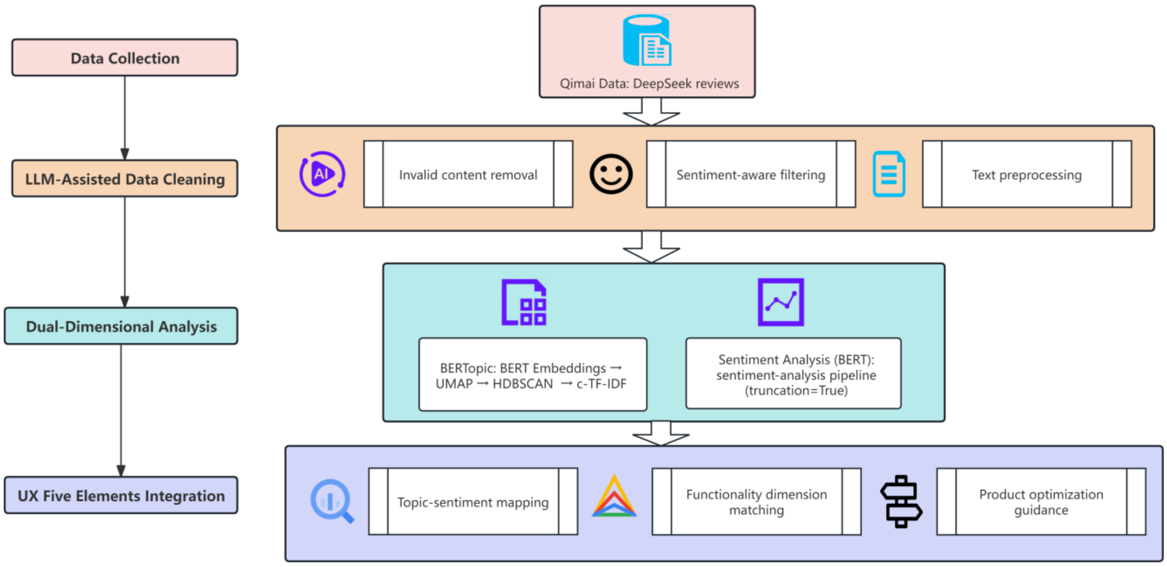

Data Collection
LLM-Assisted Data Cleaning
Dual-Dimensional Analysis
UX Five Elements Integration
Qimai Data: DeepSeek reviews
Invalid content removal
Sentiment-aware filtering
Text preprocessing
BERTopic: BERT Embeddings → UMAP → HDBSCAN → c-TF-IDF
Sentiment Analysis (BERT): sentiment-analysis pipeline (truncation=True)
Topic-sentiment mapping
Functionality dimension matching
Product optimization guidance

Figure 1. The technical roadmap of this research.

Data collection and cleaning

The dataset was obtained from Qimai Data (https://www.qimai.cn), a major Chinese mobile app analytics platform that aggregates reviews from the Apple App Store, Google Play, and domestic Android markets. Reviews of DeepSeek were collected from January 11, 2025 (official release) to April 21, 2025, covering the initial adoption stage. A total of 8,469 reviews were gathered, and after preprocessing, 5,362 valid entries remained, ensuring adequate scale, platform diversity, and temporal continuity.

To ensure analytical validity, a two-stage data cleaning procedure was implemented, combining manual screening with automated assistance. In the first stage, preliminary suggestions (‘*retain*’ or ‘*remove*’) for each review were generated using the GPT-4o model based on semantic relevance, emotional content, and informational value. This automated step was applied solely as an aid to manual screening, and all final inclusion or exclusion decisions were made by the research team. In the second stage, a manual verification procedure was conducted to review borderline cases and confirm deletion decisions, preventing the accidental removal of substantive feedback. Following this two-stage process, only invalid entries were permanently removed.

The final procedure included three operations:

(1) Invalid review removal: duplicates, advertisements, meaningless strings, and platform-marked deletions were excluded;

(2) Sentiment-aware filtering: reviews with identifiable positive or negative sentiment were retained, while nonsensical fragments were eliminated;

(3) Standard preprocessing: tokenisation via jieba, stop-word filtering (HIT list and domain terms), and normalisation were applied to prepare the corpus for topic modelling.

This hybrid approach, in which the large language model (LLM) served only as a preliminary aid, enhanced both the efficiency and the reliability of unstructured text handling. Table 1 summarises the screening criteria and representative examples.

Element	Definition	Processing Standard	Output Result Example
Review Content	The actual text of user reviews	Retain or remove based on semantic relevance judgment followed by manual verification	[1] Retain
Emotional Expression	Emotional tendency recognition	Reviews that clearly contain positive/negative emotional vocabulary	[1] Retain
Evaluation Suggestion	Extraction of specific function feedback	Reviews involving the evaluation of functions, scenarios, or characteristics	[1] Retain
Invalid Review	If the review contains the field ‘*This review has been deleted*’	Labelled by LLM as ‘NONE’ and confirmed in manual review	NONE
Invalid Review	Filtering of advertisements or irrelevant content	Removed after LLM labelling and manual confirmation	[2] Remove

Table 1. Review screening scheme with LLM labelling and manual verification.

This method effectively improves the processing efficiency of unstructured text data by drawing on cutting-edge research results in large model data cleaning (Castro et al., 2024; Li et al., 2024). Finally, a total of 5,362 valid pieces of data determined to be retained were obtained for the next step of analysis.

Data analysis

Themes were extracted using BERTopic (Grootendorst, 2022), which integrates BERT embeddings, uniform manifold approximation and projection (UMAP) for dimensionality reduction, and hierarchical density-based spatial clustering of applications with noise (HDBSCAN) for clustering. Each review was converted into a semantic embedding by a pre-trained BERT model, followed by UMAP reduction (n_neighbors = 15, min_dist = 0.001). HDBSCAN, with min_cluster_size = 15, partitioned clusters while handling noise. Class-based TF-IDF was then used to extract topic keywords, with a customised tokeniser for Chinese text. The model generated 34 coherent topics, while unclustered reviews were labelled as noise.

Sentiment labels were mainly determined from the star ratings provided by each user review. Reviews with 4_5 stars were considered positive, 3 stars as neutral, and 1_2 stars as negative. In addition, a sentiment analysis pipeline based on the Hugging Face Transformers library (https://huggingface.co/docs/transformers) was applied. The pipeline implementation, with truncation enabled, ensured efficiency in long-text processing. Results were structured in tabular form and visualised through charts. Sentiment distributions across topics were further examined to reveal emotional variations among functional themes.

Analytical results were aligned with Garrett's five elements of user experience framework (Garrett, 2010) to generate actionable insights. Topic–sentiment mappings were associated with dimensions such as functionality and usability, clarifying which features elicited positive endorsement or negative frustration. For instance, topics receiving concentrated negative sentiment were linked to functionality and usability, signalling areas in need of optimisation.

This integrated framework provides a systematic approach for understanding user needs and emotions, offering practical guidance for evidence-based product refinement.

Findings

This chapter presents user feedback data on the DeepSeek APP, illustrating key results from BERTopic clustering and sentiment analysis with specific metrics and visualisations. BERTopic clustering clarifies the distribution and characteristics of user feedback topics, and sentiment analysis further reveals user emotional tendencies and topic-specific differences. By further integrating the results of topic clustering and sentiment analysis with the five-element theory of user experience, the core issues in user feedback are identified.

Topic clustering results

BERTopic clustering was applied to the user reviews, yielding two primary sets of results. The analysis first identifies 34 specific topics, detailing their keywords, data volume, and content focus to pinpoint specific user concerns. Secondly, a high-level perspective is provided by hierarchically clustering the 34 topics into nine major categories based on semantic similarity. This integration, supported by distance visualisation, illustrates the overall distribution of user concerns.

BERTopic core topic analysis

BERTopic clustering identified 34 topics, each represented by five keywords. Figure 2 summarises each topic's name, keyword list (Chinese and English), and sample size, derived from keyword connotations and cluster review counts.

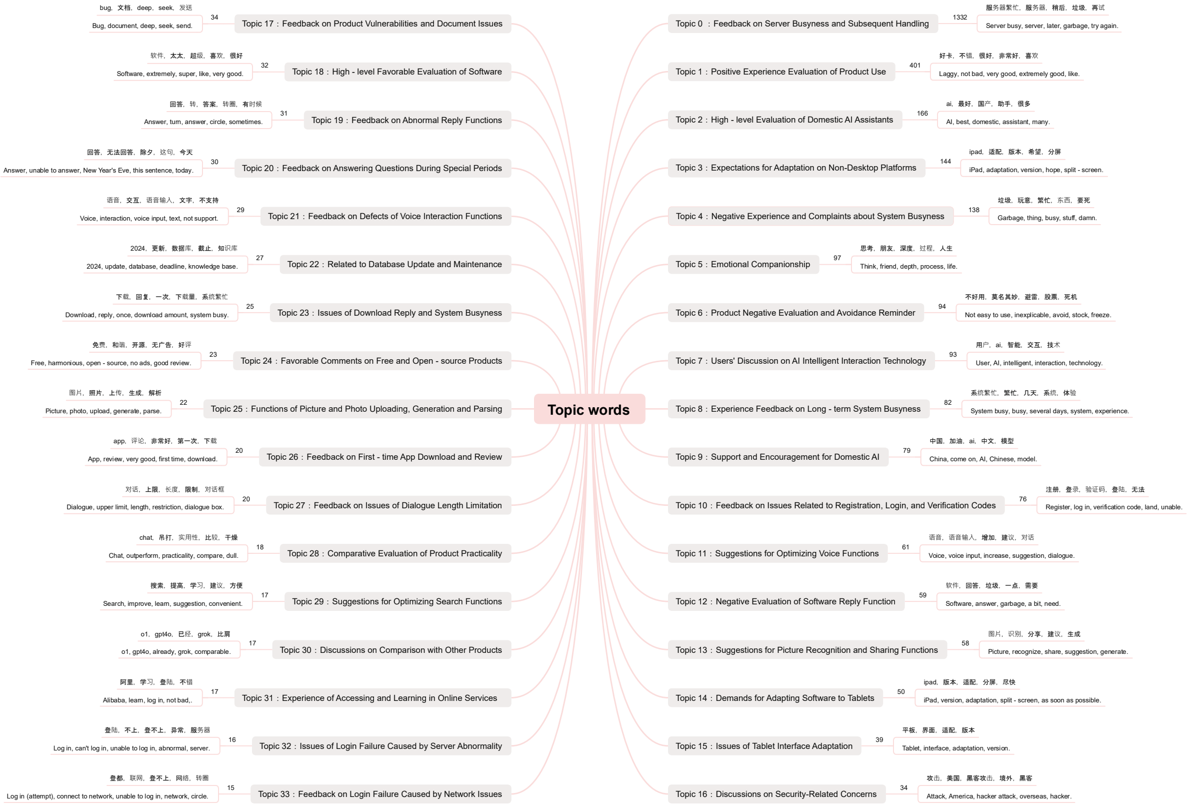

Topic words
Topic 0 ：Feedback on Server Busyness and Subsequent Handling
1332
服务器繁忙，服务器，稍后，垃圾，再试
Server busy, server, later, garbage, try again.
Topic 1：Positive Experience Evaluation of Product Use
401
好卡，不错，很好，非常好，喜欢
Laggy, not bad, very good, extremely good, like.
Topic 2：High - level Evaluation of Domestic AI Assistants
166
ai，最好，国产，助手，很多
AI, best, domestic, assistant, many.
Topic 3：Expectations for Adaptation on Non-Desktop Platforms
144
ipad，适配，版本，希望，分屏
iPad, adaptation, version, hope, split - screen.
Topic 4：Negative Experience and Complaints about System Busyness
138
垃圾，玩意，繁忙，东西，要死
Garbage, thing, busy, stuff, damn.
Topic 5：Emotional Companionship
97
思考，朋友，深度，过程，人生
Think, friend, depth, process, life.
Topic 6：Product Negative Evaluation and Avoidance Reminder
94
不好用，莫名其妙，避雷，股票，死机
Not easy to use, inexplicable, avoid, stock, freeze.
Topic 7：Users' Discussion on AI Intelligent Interaction Technology
93
用户，ai，智能，交互，技术
User, AI, intelligent, interaction, technology.
Topic 8：Experience Feedback on Long - term System Busyness
82
系统繁忙，繁忙，几天，系统，体验
System busy, busy, several days, system, experience.
Topic 9：Support and Encouragement for Domestic AI
79
中国，加油，ai，中文，模型
China, come on, AI, Chinese, model.
Topic 10：Feedback on Issues Related to Registration, Login, and Verification Codes
76
注册，登录，验证码，登陆，无法
Register, log in, verification code, land, unable.
Topic 11：Suggestions for Optimizing Voice Functions
61
语音，语音输入，增加，建议，对话
Voice, voice input, increase, suggestion, dialogue.
Topic 12：Negative Evaluation of Software Reply Function
59
软件，回答，垃圾，一点，需要
Software, answer, garbage, a bit, need.
Topic 13：Suggestions for Picture Recognition and Sharing Functions
58
图片，识别，分享，建议，生成
Picture, recognize, share, suggestion, generate.
Topic 14：Demands for Adapting Software to Tablets
50
ipad，版本，适配，分屏，尽快
iPad, version, adaptation, split - screen, as soon as possible.
Topic 15：Issues of Tablet Interface Adaptation
39
平板，界面，适配，版本
Tablet, interface, adaptation, version.
Topic 16：Discussions on Security-Related Concerns
34
攻击，美国，黑客攻击，境外，黑客
Attack, America, hacker attack, overseas, hacker.
Topic 17：Feedback on Product Vulnerabilities and Document Issues
34
bug，文档，deep，seek，发送
Bug, document, deep, seek, send.
Topic 18：High - level Favorable Evaluation of Software
32
软件，太太，超级，喜欢，很好
Software, extremely, super, like, very good.
Topic 19：Feedback on Abnormal Reply Functions
31
回答，转，答案，转圈，有时候
Answer, turn, answer, circle, sometimes.
Topic 20：Feedback on Answering Questions During Special Periods
30
回答，无法回答，除夕，这句，今天
Answer, unable to answer, New Year's Eve, this sentence, today.
Topic 21：Feedback on Defects of Voice Interaction Functions
29
语音，交互，语音输入，文字，不支持
Voice, interaction, voice input, text, not support.
Topic 22：Related to Database Update and Maintenance
27
2024，更新，数据库，截止，知识库
2024, update, database, deadline, knowledge base.
Topic 23：Issues of Download Reply and System Busyness
25
下载，回复，一次，下载量，系统繁忙
Download, reply, once, download amount, system busy.
Topic 24：Favorable Comments on Free and Open - source Products
23
免费，和谐，开源，无广告，好评
Free, harmonious, open - source, no ads, good review.
Topic 25：Functions of Picture and Photo Uploading, Generation and Parsing
22
图片，照片，上传，生成，解析
Picture, photo, upload, generate, parse.
Topic 26：Feedback on First - time App Download and Review
20
app，评论，非常好，第一次，下载
App, review, very good, first time, download.
Topic 27：Feedback on Issues of Dialogue Length Limitation
20
对话，上限，长度，限制，对话框
Dialogue, upper limit, length, restriction, dialogue box.
Topic 28：Comparative Evaluation of Product Practicality
18
chat，吊打，实用性，比较，干燥
Chat, outperform, practicality, compare, dull.
Topic 29：Suggestions for Optimizing Search Functions
17
搜索，提高，学习，建议，方便
Search, improve, learn, suggestion, convenient.
Topic 30：Discussions on Comparison with Other Products
17
o1，gpt4o，已经，grok，比肩
o1, gpt4o, already, grok, comparable.
Topic 31：Experience of Accessing and Learning in Online Services
17
阿里，学习，登陆，不错
Alibaba, learn, log in, not bad,.
Topic 32：Issues of Login Failure Caused by Server Abnormality
16
登陆，不上，登不上，异常，服务器
Log in, can't log in, unable to log in, abnormal, server.
Topic 33：Feedback on Login Failure Caused by Network Issues
15
登都，联网，登不上，网络，转圈
Log in (attempt), connect to network, unable to log in, network, circle.

Figure 2. Graphical map of topic words in Chinese and English.

Topic 0 (largest cluster, 1332 reviews) focuses on server busyness and handling, revealing technical failures and user frustration—confirming server stability as a core bottleneck. Topic 1 (401 reviews) reflects positive evaluations of basic interactions, serving as a benchmark for retained advantages. Topic 2 (166 reviews) and Topic 9 (79 reviews) show strong approval for domestic AI, supporting DeepSeek's differentiated positioning as a Chinese AI product.

For device adaptation, Topic 3 (iPad expectations), Topic 14 (iPad demands), and Topic 15 (tablet interface issues) collectively account for 233 reviews, highlighting user demand for smooth cross-device use. For advanced functions, Topic 11 (voice optimisation) and Topic 13 (image recognition/sharing) emphasise need for voice accuracy, response speed, and simplified image operations—guiding improvements from '*usable*' to '*user-friendly*.' Technical defects, though in smaller clusters, require high priority: Topic 17 (bugs/documents), Topic 21 (voice interaction defects), Topic 32 (server-related login failure), Topic 19 (abnormal replies), and Topic 20 (special-period answering issues) all impact core functionality.

In summary, optimisation priorities are: (1) improve server stability (Topic 0) and fix critical defects (Topics 17, 21, 32); (2) optimise multi-device adaptation (Topics 3, 14, 15) and advanced functions (Topics 11, 13); (3) maintain advantages in basic interactions (Topic 1) and domestic AI recognition (Topics 2, 9).

BERTopic thematic clustering analysis

The BERTopic hierarchical clustering results reveal the distribution of user reviews across multiple themes and the underlying structural relationships. As shown in the hierarchical clustering (Figure 3), user feedback is highly diverse, with themes exhibiting both distinct boundaries and interconnections. In the visualisation, each colour represents a topic cluster formed through hierarchical merging, indicating groups of semantically related topics. Based on detailed examination, these themes were consolidated into nine major categories that capture the breadth of user concerns and experiences.

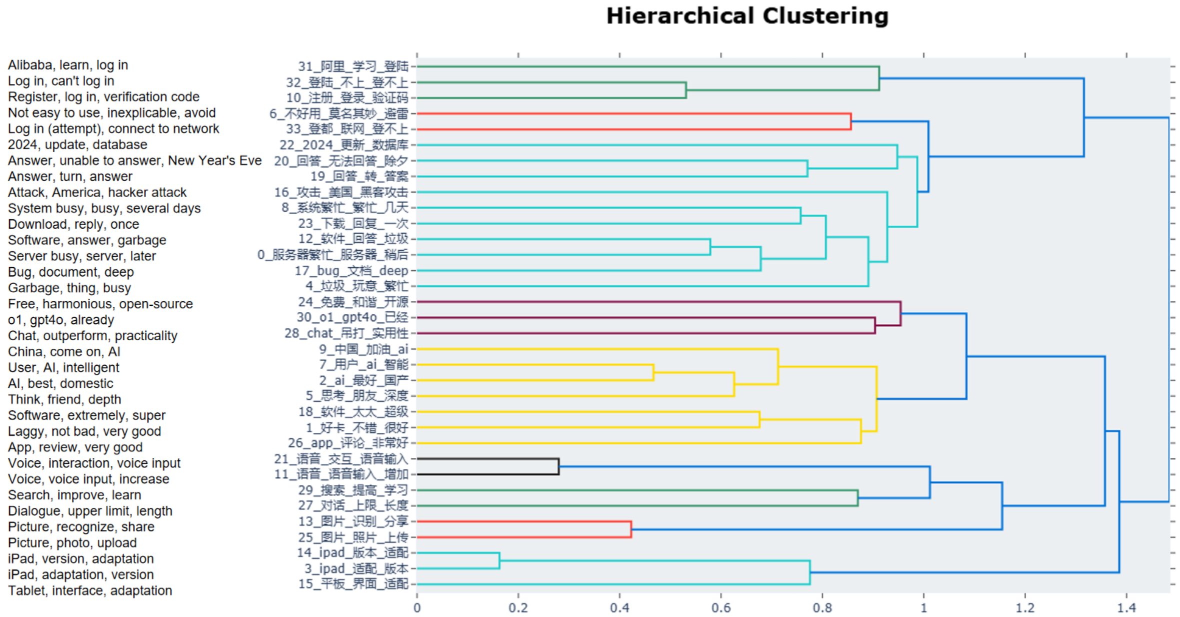


Figure 3. Thematic hierarchical clustering diagram.

(1) Technical and server issues

This category focuses on recurring technical problems, including server failures, overload, and software vulnerabilities. Topic 0, the largest cluster (1,332 entries), centres on keywords like '*server busy*' and '*try again*,' reflecting widespread user frustration with server instability; Topic 8 highlights the impact of prolonged system busyness, while Topic 17 relates to bugs and document-related failures. These findings confirm that technical stability is a core factor shaping user experience.

(2) User evaluation and feedback

User sentiment shows a clear polarised pattern. Topic 1 (401 entries) contains positive evaluations such as '*very good*' and '*like*,' while Topic 4 and Topic 6 include negative descriptors like '*garbage*' and '*not easy to use*.' This contrast between positive and negative feedback provides direct guidance for targeted product optimisation.

(3) Device adaptation and functions

This category centres on adaptation needs for iPads and tablets, with a particular focus on requests for split-screen support and interface refinement. Topic 3 (144 entries) is the main carrier of such feedback, while Topic 14 and Topic 15 further supplement issues related to tablet interface adaptation. This indicates that cross-device compatibility is crucial for improving user satisfaction.

(4) AI technology and applications

Several themes reflect strong user recognition of domestic generative AI: Topic 2 praises the app's AI capabilities, while Topic 7 and Topic 9 involve intelligent interaction and national pride in Chinese AI models. It is evident that user expectations for the product extend beyond basic functionality to include deeper technological and cultural value.

(5) Account and login

Problems in this category focus on verification code abnormalities and login failures, often linked to server instability, mainly involving Topic 10, Topic 32, and Topic 33. This phenomenon underscores the critical role of infrastructure in ensuring seamless user access.

(6) Image functions

User demands for image recognition, sharing, and generation features are primarily reflected in Topic 13 and Topic 25. As image-based interaction becomes more prevalent, strengthening these functions has become a necessary direction for improving product usability.

(7) Product features and security

Topic 24 reflects user appreciation for the app's free, open-source, and ad-free features, while Topic 16 concerns security risks such as hacking and overseas attacks. Together, they reveal that affordability and trustworthiness are the dual keys to sustaining user engagement.

(8) Functional abnormalities and special scenarios

Topic 19 discusses anomalies in the answering function, while Topic 20 focuses on service failures during peak periods. This shows that stability in high-demand scenarios is as crucial as daily reliability.

(9) Download and interaction

Topic 23 highlights inefficiencies in the download process, while Topic 27 addresses dialogue length limitations. Such feedback emphasises the need to further optimise interaction smoothness and reduce operational friction for users.

Overall, the nine thematic categories span technical infrastructure, user sentiment, device adaptation, AI applications, account access, image processing, product features, functional stability, and interaction design. They reveal both persistent technical pain points and sources of user appreciation, such as open-source availability and domestic AI innovation. Overall, these topics offer comprehensive feedback on user experience, providing referenceable ideas for goal optimisation and user-centred product development.

Results of sentiment analysis

The results of sentiment analysis are presented through two visualisations: Figure 4 illustrates the overall distribution of review sentiments, while Figure 5 depicts the sentiment tendency across different topics.

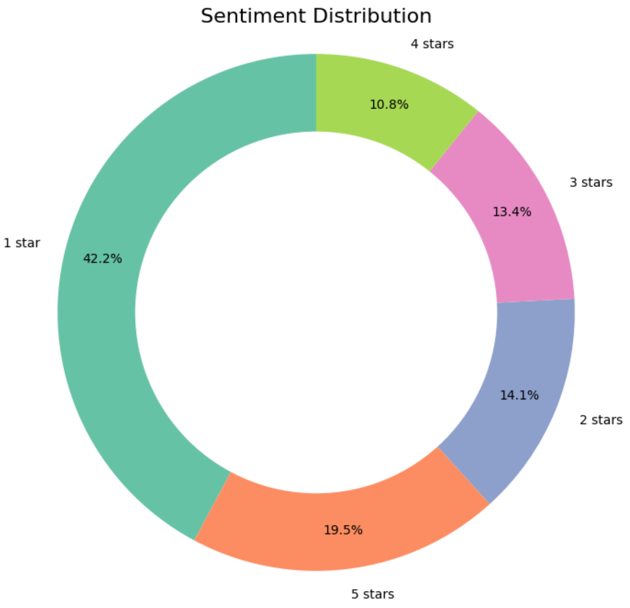


Figure 4. Distribution chart of review sentiments.

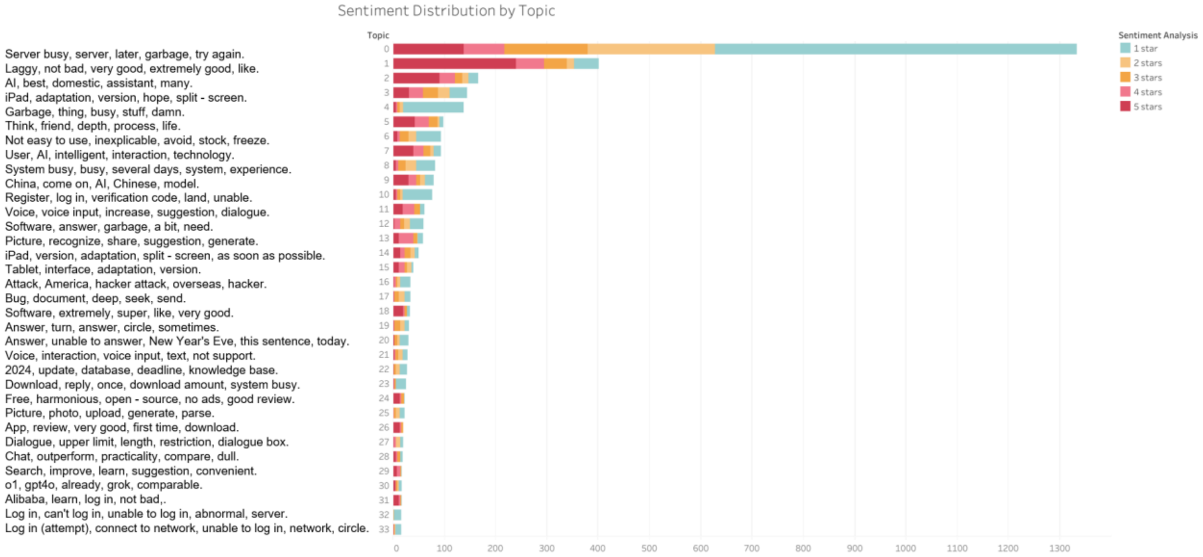


Figure 5. Visualisation of topic-sentiment distribution.

The overall sentiment distribution (Figure 4) reveals a pronounced tendency toward negative evaluations. Notably, 1-star reviews constitute 42.2% of the total, suggesting that nearly half of users express dissatisfaction and perceive substantial shortcomings in product experience. By contrast, highly positive feedback is limited: 5-star reviews account for 19.5% and 4-star reviews 10.8%. Neutral (3-star) and moderately negative (2-star) reviews represent 13.4% and 14.1% respectively. In sum, only 30.3% of reviews are positive (4-star and 5-star combined), underscoring the dominance of negative sentiment and the urgent need for product refinement.

The sentiment distribution across topics (Figure 5) further highlights differences in emotional responses. Topics centred on technical and server instability show an overwhelming negative sentiment, with 1-star ratings significantly outnumbering other star levels. Specifically, Topic 0 focuses on feedback regarding server busyness and subsequent handling, while Topic 8 centres on feedback about long-term system busyness and user experience. The consistent negative tendency of these two topics indicates that recurring system failures and unstable infrastructure directly undermine users' trust in the product.

Positive evaluations of user-endorsed topics are concentrated in two scenarios, with a notably higher proportion of 4-star and 5-star reviews—reflecting users' recognition of the product's usability and accessibility. One scenario relates to product usage experience, covered by Topic 1, which mainly carries users' positive evaluations of product use. The other relates to product attributes, focused on Topic 24, which gathers users' favourable reviews on free and open-source products. These two types of topics together form the core source of positive sentiment. Topics related to domestic AI development also show a dominant high-star rating tendency, consistent with the positive inclination of user-endorsed topics but with a more specific focus. Among them, Topic 2 covers high-level evaluations of domestic AI assistants, and Topic 9 focuses on support and encouragement for domestic AI. The high-star ratings of these two topics directly reflect users' emotional identification with local AI innovation. Topics involving functional anomalies and situational use exhibit a clear negative tendency, with a prominent proportion of 1-star and 2-star reviews. The negative sentiment of these topics mainly stems from unmet functional expectations: Topic 19 addresses feedback on abnormal reply functions, and Topic 20 relates to feedback on answering questions during special periods. Both topics confirm that functional defects directly translate into users' negative affect.

In summary, four core conclusions can be drawn: 1) Technical stability is the primary inducement of negative sentiment; 2) Positive sentiment mainly stems from product performance, cost-effectiveness, and emotional identification with domestic AI; 3) Functional anomalies in specific scenarios significantly intensify negative perceptions; 4) Sentiment tendencies are highly associated with user experience, which can provide targeted basis for subsequent optimization.

Analysis of the five elements of user experience

The five elements of user experience, as a crucial framework for evaluating a product's alignment with user needs, offers a comprehensive diagnosis of the product experience across its strategy, scope, structure, framework, and surface planes. Based on the preceding analysis, Figure 6 will be used to explain the current state of the product experience at each plane, providing a clear direction for the formulation of optimisation strategies.

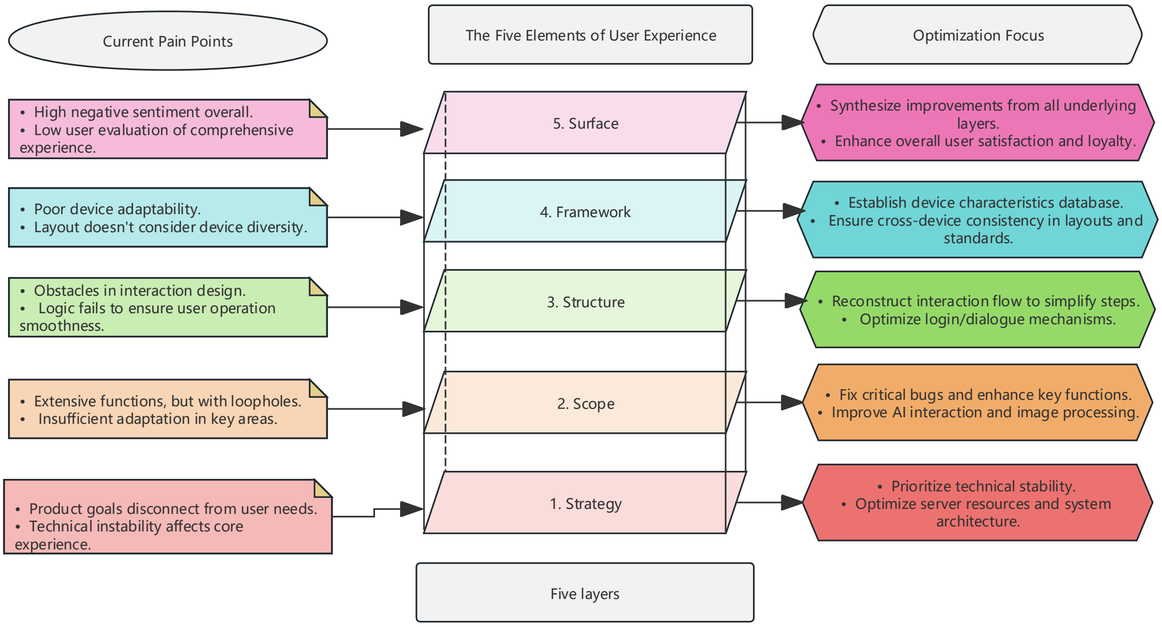
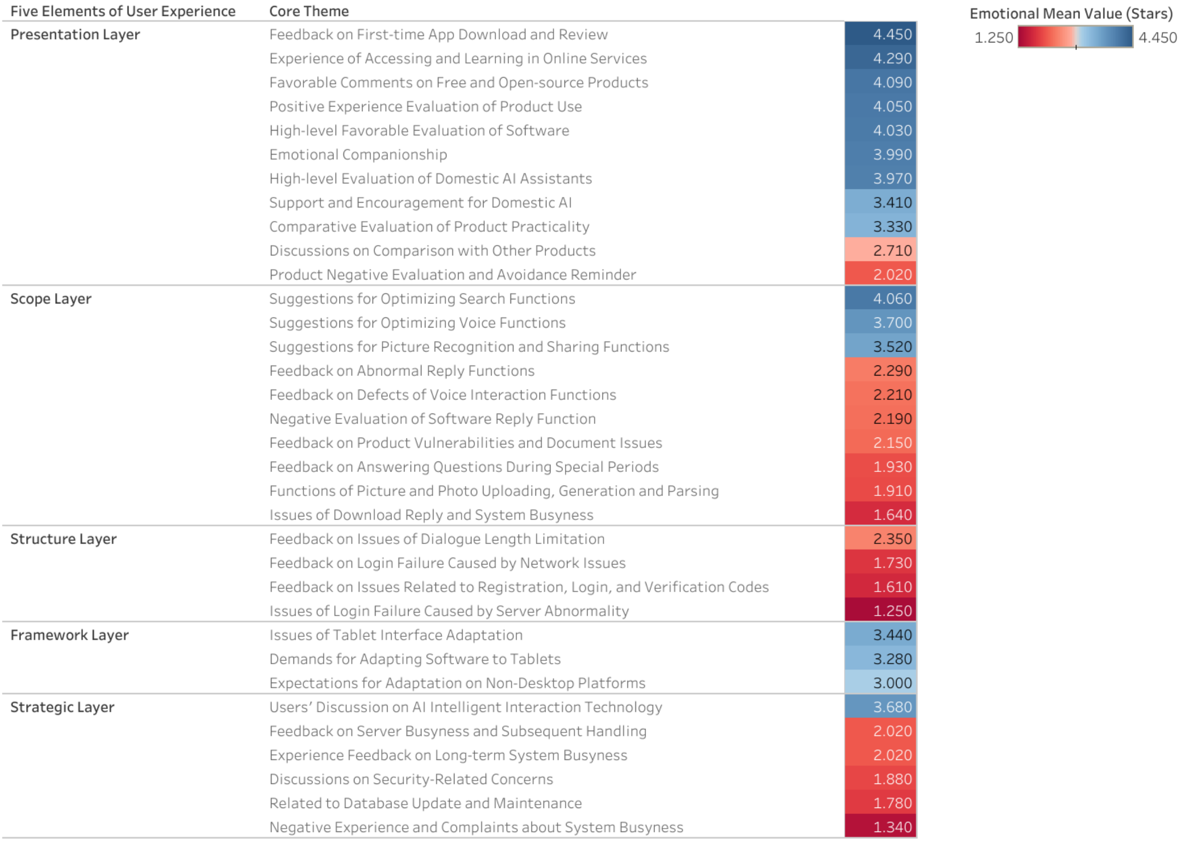

Current Pain Points
The Five Elements of User Experience
Optimization Focus
• High negative sentiment overall.
• Low user evaluation of comprehensive experience.
5. Surface
• Synthesize improvements from all underlying layers.
• Enhance overall user satisfaction and loyalty.
• Poor device adaptability.
• Layout doesn't consider device diversity.
4. Framework
• Establish device characteristics database.
• Ensure cross-device consistency in layouts and standards.
• Obstacles in interaction design.
• Logic fails to ensure user operation smoothness.
3. Structure
• Reconstruct interaction flow to simplify steps.
• Optimize login/dialogue mechanisms.
• Extensive functions, but with loopholes.
• Insufficient adaptation in key areas.
2. Scope
• Fix critical bugs and enhance key functions.
• Improve AI interaction and image processing.
• Product goals disconnect from user needs.
• Technical instability affects core experience.
1. Strategy
• Prioritize technical stability.
• Optimize server resources and system architecture.
Five layers

Figure 6. Analysis of the five elements of user experience.



Five Elements of User Experience	Core Theme	
Presentation Layer	Feedback on First-time App Download and Review	4.450
	Experience of Accessing and Learning in Online Services	4.290
	Favorable Comments on Free and Open-source Products	4.090
	Positive Experience Evaluation of Product Use	4.050
	High-level Favorable Evaluation of Software	4.030
	Emotional Companionship	3.990
	High-level Evaluation of Domestic AI Assistants	3.970
	Support and Encouragement for Domestic AI	3.410
	Comparative Evaluation of Product Practicality	3.330
	Discussions on Comparison with Other Products	2.710
	Product Negative Evaluation and Avoidance Reminder	2.020
Scope Layer	Suggestions for Optimizing Search Functions	4.060
	Suggestions for Optimizing Voice Functions	3.700
	Suggestions for Picture Recognition and Sharing Functions	3.520
	Feedback on Abnormal Reply Functions	2.290
	Feedback on Defects of Voice Interaction Functions	2.210
	Negative Evaluation of Software Reply Function	2.190
	Feedback on Product Vulnerabilities and Document Issues	2.150
	Feedback on Answering Questions During Special Periods	1.930
	Functions of Picture and Photo Uploading, Generation and Parsing	1.910
	Issues of Download Reply and System Busyness	1.640
Structure Layer	Feedback on Issues of Dialogue Length Limitation	2.350
	Feedback on Login Failure Caused by Network Issues	1.730
	Feedback on Issues Related to Registration, Login, and Verification Codes	1.610
	Issues of Login Failure Caused by Server Abnormality	1.250
Framework Layer	Issues of Tablet Interface Adaptation	3.440
	Demands for Adapting Software to Tablets	3.280
	Expectations for Adaptation on Non-Desktop Platforms	3.000
Strategic Layer	Users' Discussion on AI Intelligent Interaction Technology	3.680
	Feedback on Server Busyness and Subsequent Handling	2.020
	Experience Feedback on Long-term System Busyness	2.020
	Discussions on Security-Related Concerns	1.880
	Related to Database Update and Maintenance	1.780
	Negative Experience and Complaints about System Busyness	1.340

Figure 7. Heatmap of the five elements of user experience and topic-sentiment mean values.

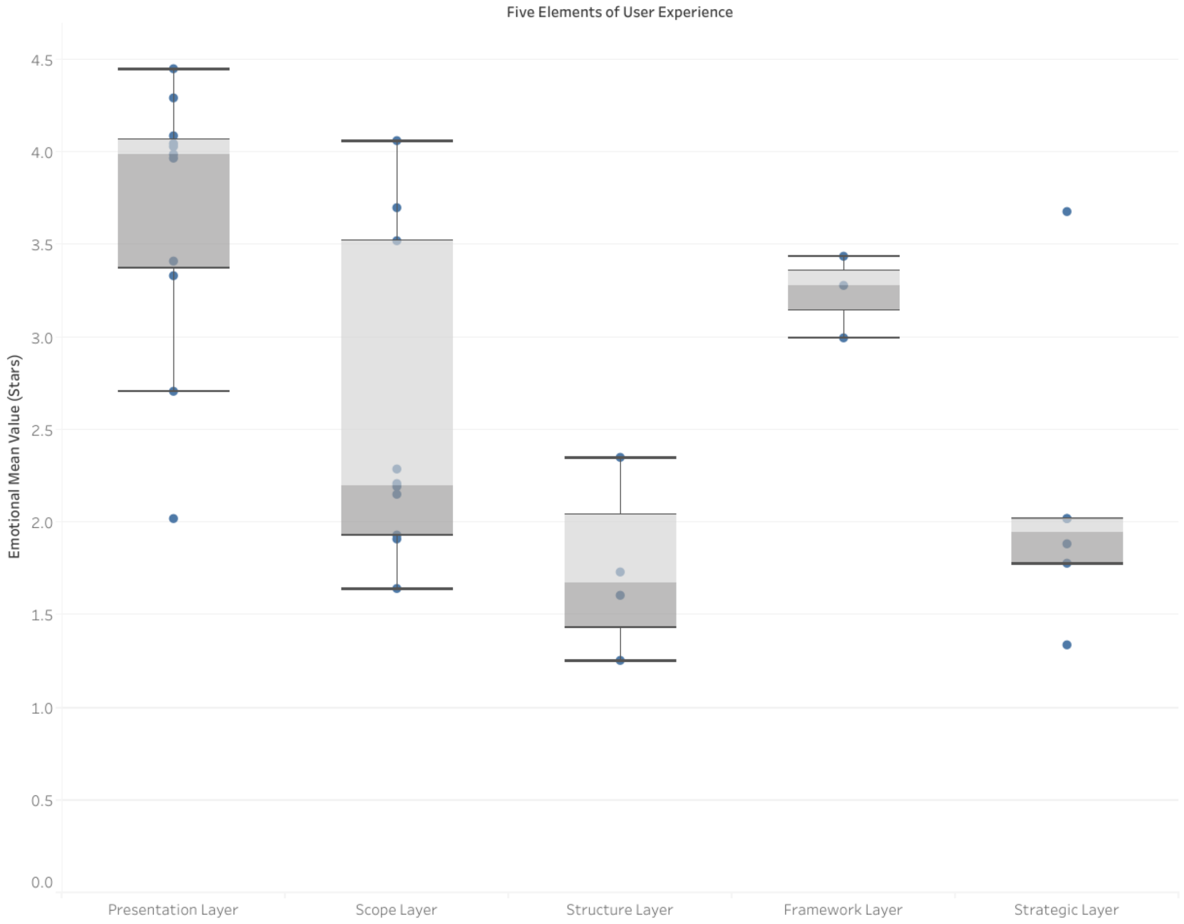


Figure 8. Box plot of sentiment mean values for the five elements of user experience.

By grouping and aggregating the star ratings for all reviews within each topic, a mean sentiment value for each topic was calculated (Mean Sentiment = sum of review star ratings (1–5, 5 = most positive) within a topic/number of reviews), resulting in a heatmap and a box plot.

Figure 7 and Figure 8 respectively present the hierarchical results of the five elements based on BERTopic topic modelling and the calculation of the mean sentiment, as well as the sentiment distribution characteristics. The heat map reveals the differences in the mean sentiment of user evaluations under different topics: some topics in the performance plane and the scope plane have relatively high means, reflecting that users have certain positive feedback in the initial contact and the use of some functions; however, topics related to system stability, login interaction, download and reply functions are generally in the low mean range, indicating a significant concentration of negative sentiments.

The box plot further shows the dispersion of sentiment distribution within different planes, among which the user sentiment in the performance plane is the most positive, with feedback generally concentrated and the highest satisfaction. The mean sentiment of the framework plane is relatively high and concentrated, indicating that user feedback is relatively positive. The sentiment feedback in the scope plane has the greatest difference, suggesting that there are both extremely good and extremely poor topics within this plane. While the strategy plane and the structure plane show a relatively concentrated but overall lower evaluation trend.

Discussion

This study provides a structured understanding of DeepSeek's user experience by integrating topic analysis from BERTopic, sentiment assessment, and Garrett's five elements of user experience (Garrett, 2010). Analysing approximately 8,000 Chinese user reviews, the study identifies latent topics and emotional responses across strategic, functional, structural, framework, and surface aspects. Key concerns include technical stability, functional completeness, and interaction flow, while positive feedback highlights AI innovation, accessibility, and emotional recognition. These insights offer a comprehensive view of user satisfaction, guiding improvements for DeepSeek and similar AI applications.

Analysis of latent topics in DeepSeek's Chinese user reviews

The topic analysis identified 34 core topics, which were further synthesised into nine categories covering technical and server issues, user evaluations, device adaptation, and functional features. These categories highlight the key areas of user concern, with negative sentiment concentrated on server overload, unreliable performance, and functional vulnerabilities, whereas positive feedback emphasised the application's affordability, domestic AI innovation, and perceived utility. Mapping these topics onto Garrett's five elements of user experience provides a structured lens for interpreting the patterns of user experience.

Strategic plane: The strategic dimension reflecting the consistency between product goals and users' core expectations is significantly challenged by technological instability. One-star negative reviews (42.2%) are mainly focused on server overload and system crashes, indicating that the primary demand for reliability from users has not been met. Notably, time analysis shows that these complaints were most prevalent during the early release stage. This was due to the provision of a high-performance, free, and open-source domestic AI model, which led to an unexpected influx of a massive number of users and developers to DeepSeek, causing server overload within a short period. Studies have shown that system reliability directly affects users' trust formation and technical acceptance (Thorne, 2024). Moreover, this concentration of negative feedback on server overload can also be explained from a psychological perspective. According to the theory of negativity bias, negative experiences are more likely to capture attention and trigger expressive behaviour (Baumeister et al., 2001). Recent research has further revealed a tendency towards polarity self-selection in online reviews, whereby users who experience extreme events are more likely to post evaluations, which can result in a concentration of negative feedback (Schoenmueller et al., 2020). These findings suggest that technological stability should be prioritised as a strategic goal, as failures at this plane can spread negative perceptions throughout downstream user experience elements.

Scope plane: The product's functional coverage is extensive, but critical gaps can undermine its effectiveness. User feedback has highlighted issues in AI interaction capabilities, image processing, and device compatibility, particularly in the iPad and tablet environments. Complaints surrounding these functions indicate that incomplete or inconsistent functional implementation reduces perceived usefulness. These studies emphasise that functional reliability and comprehensive functional delivery are crucial for maintaining user satisfaction in AI applications. Therefore, scope optimisation should focus on addressing key vulnerabilities, enhancing multi-device compatibility, expanding AI and multimedia capabilities, and ensuring both the breadth and depth of functionality.

Structural plane: Interaction design and workflow impede smooth operation. Users highlighted the friction in account registration and login (e.g., failed verification codes) and the limitations on dialogue length, which collectively reduce operational fluency. Research shows that although large language models like ChatGPT have outstanding question-answering capabilities, if users cannot engage in in-depth discussions over long sections and multiple rounds, the completeness and depth of the information they obtain will be affected, thereby reducing their overall satisfaction with the system (Shen et al., 2023). To mitigate these effects, structural optimisation should streamline user workflows, simplify the authentication process, and enhance dialogue interaction mechanisms, thereby reducing operational friction and improving perceived usability.

Framework plane: Interface adaptability across devices remains inconsistent. reviews highlighted display and layout issues on the iPad and other tablet devices, indicating that cross platform interface design is not ideal. By leveraging device specifications and usage pattern databases to achieve device specific layout optimisation, interface consistency can be enhanced, and functional and aesthetic experiences can be maintained across platforms.

Surface plane: Issues in terms of strategy, scope, structure, and framework are manifested at the surface plane as overall user sentiment. The predominance of negative reviews highlights that the overall perception of the product is hindered by accumulated defects. Positive emotions are concentrated on the domestic development of artificial intelligence capabilities and the accessibility of open source, indicating that when the functional and interaction layers are adequately addressed, technological innovation and value perception can meaningfully enhance satisfaction. This further indicates that the user acceptance results of artificial intelligence technology are jointly determined by reliability, perceived usefulness, and perceived ease of use (Choung et al., 2023).

In summary, deficiencies at the foundational plane (strategy and scope) propagate through interaction and perception planes, creating a hierarchical dependency in UX elements that ultimately affects overall user satisfaction.

Factors behind user satisfaction and dissatisfaction

The sentiment analysis provides a nuanced understanding of how users perceive DeepSeek. Negative sentiment clusters around server instability, system anomalies, and functional gaps, reinforcing user dissatisfaction at multiple UX planes. Positive feedback primarily concerns functional success, emotional recognition (such as AI companionship), and alignment with domestic AI expectations. This polarity underscores the interplay between technical performance, functional completeness, and perceived value in shaping overall user experience.

Integrating sentiment with the five elements framework reveals systemic misalignments: strategic goals diverge from user stability expectations; scope-related gaps indicate incomplete functionality or adaptation issues; structural inefficiencies manifest in cumbersome registration and dialogue limitations; framework inconsistencies affect cross-device interface experience; and surface-plane sentiment reflects the cumulative effect of these issues. Such integration demonstrates the hierarchical and interconnected nature of user experience elements in generative AI applications.

Based on these insights, phased suggestions for improving user experience were put forward. At the strategic plane, prioritising technological stability and server reliability is essential. Scope optimisation should address key functional gaps, improve multi-device compatibility, and ensure consistency in AI and multimedia capabilities. Structural improvements include streamlining user workflows, simplifying authentication, and enhancing dialogue mechanisms. Framework-plane recommendations involve responsive design and interface adaptation to maintain consistency across devices. At the surface plane, fostering positive emotional engagement through functional reliability and highlighting AI innovation can enhance perceived value.

Suggestions for optimising user experience

Based on the topic and sentiment analysis of DeepSeek user comments, systematic optimisation recommendations can be proposed to enhance overall user experience.

At the strategic plane, users are most concerned with server stability and reliability, with frequent reports of crashes, delays, and overloads undermining trust. Implementing load balancing, automated monitoring, rapid response protocols, and enhanced performance testing can ensure service continuity and lay a stable foundation for further improvements.

At the scope plane, feedback highlights missing core functionalities, inconsistent AI interaction, and limited device compatibility, especially on tablets. Refining key features, standardising AI behaviour, optimising multi-device adaptation, and expanding functional coverage can deliver a more cohesive and valuable experience across scenarios.

At the structural plane, registration, login, and conversation interactions create friction, such as failed verification codes or restricted dialogue lengths. Streamlining these processes, improving dialogue mechanisms, and incorporating guidance and error tolerant design can reduce friction and enhance usability.

At the framework plane, inconsistent interfaces across devices affect perception and visual continuity. Responsive design, device specific layouts, and unified interface standards can improve cross-platform consistency and aesthetic experience.

At the surface plane, sentiment reflects the cumulative effect of underlying planes. Positive emotions relate to innovation, open-source advantages, and domestic AI features, while negative sentiment stems from unresolved issues. Emphasising technological value and domestic AI strengths, and establishing dynamic feedback tracking, can reinforce satisfaction and create a continuous improvement loop.

Through these strategies, DeepSeek can improve interaction quality, functional consistency, and perceived value while maintaining core stability, transitioning from reactive problem-solving to proactive value creation.

Conclusion

This study systematically examined user reviews of the DeepSeek APP by integrating BERTopic topic modelling with sentiment analysis. The analysis identified 34 core topics, which were further synthesised into nine categories, covering technical and server issues, user evaluations, device adaptation, and functionality. Sentiment analysis revealed that 42.2% of reviews expressed negative experiences, largely related to server instability, system failures, and functional vulnerabilities, while 30.3% of positive reviews highlighted the app's practicality and its advantages as a free, open-source domestic generative AI application. An additional assessment using the five elements of user experience underscored substantial optimisation needs across all planes—strategy, scope, structure, framework, and surface.

This study follows an analytical pathway that begins with identifying user-reported problems and then moves towards diagnosing experience elements and formulating optimisation strategies. The analysis systematically processing user reviews and turning fragmented, subjective experiences into interpretable patterns of user experience. Through this method, it provides an empirical evidence base for optimising generative AI applications, rather than relying on subjective assumptions. Meanwhile, generative AI applications are treated as information systems embedded in information practices. User reviews capture not only evaluations of specific functionalities but also how users understand system capabilities, interaction mechanisms, and output quality in practical usage scenarios. Based on empirical analysis of authentic feedback from Chinese users, the findings further indicate that this methodology effectively identifies interaction and experience issues. It provides an expandable and transferable analytical pathway for understanding the use of generative AI. Moreover, these findings further underscore the value of a user-centred, data-driven approach to guiding the responsible design and deployment of generative AI in information practices.

Nevertheless, the study has limitations. It relied solely on textual review data, omitting behavioural metrics such as session duration and interaction logs that could reveal implicit needs. Methodologically, topic modelling parameters may influence topic granularity, and sentiment analysis did not account for demographic differences in user perspectives.

Future research should therefore: (1) integrate behavioural and textual data for multi-source analysis, (2) combine sentiment analysis with user profiling to capture group-specific experiences, and (3) conduct longitudinal review tracking to establish a closed-loop evaluation of '*analysis–optimisation–feedback*.' Advancing along these directions will enhance both the methodological depth and practical utility of user experience research.

Acknowledgements

This work is supported by Key Project of the National Social Science Fund of China (2024), titled 'Research on Research Support Services of University Libraries for AI for Science (AI4S)' (Grant No. 24ATQ008)

About the authors

Chenxin Zhou is a master student in School of Information Management, Wuhan University, People's Republic of China. Her research interests include the library services, research data management and user behaviour. She can be contacted at zcxmxcz@163.com

Lihong Zhou is a Professor in School of Information Management, Wuhan University, People's Republic of China. His research interests include the library services, interorganisational data sharing and medical information. He can be contacted at L.zhou@whu.edu.cn

References

Ahmed, A., Aziz, S., Khalifa, M., Shah, U., Hassan, A., Abd-Alrazaq, A., & Househ, M. (2022). Thematic analysis on user reviews for depression and anxiety chatbot apps: Machine learning approach. JMIR Formative Research, 6(3), e27654. https://doi.org/10.2196/27654

Alabduljabbar, R. (2024). User-centric AI: Evaluating the usability of generative AI applications through user reviews on app stores. PeerJ Computer Science, 10, e2421. https://doi.org/10.7717/peerj-cs.2421

Alhejji, S., Albesher, A., Wahsheh, H., & Albarrak, A. (2022). Evaluating and comparing the usability of mobile banking applications in Saudi Arabia. Information, 13(12), 559. https://doi.org/10.3390/info13120559

Alshammare, H., Alshayeb, M., & Baslyman, M. (2025). Revealing the mobile UX horizon: Exploring user experience aspects, attributes, and measurement methods-A systematic mapping study. Computer Standards & Interfaces, 94, 103999. https://doi.org/10.1016/j.csi.2025.103999

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human–AI interaction. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3290605.3300233

Baj-Rogowska, A., & Sikorski, M. (2023). Exploring the usability and user experience of social media apps through a text mining approach. Engineering Management in Production and Services, 15(1), 86-105. https://doi.org/10.2478/emj-2023-0007

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. Review of General Psychology, 5(4), 323-370. https://doi.org/10.1037/1089-2680.5.4.323

Bubaš, G., Čižmešija, A., & Kovačić, A. (2024). Development of an assessment scale for measurement of usability and user experience characteristics of Bing Chat conversational AI. Future Internet, 16(1), 4. https://doi.org/10.3390/fi16010004

Castro, A., Pinto, J., Reino, L., Pipek, P., & Capinha, C. (2024). Large language models overcome the challenges of unstructured text data in ecology. Ecological Informatics, 82, 102742. https://doi.org/10.1016/j.ecoinf.2024.102742

Chen, X., Gao, C., Chen, C., Zhang, G., & Liu, Y. (2025). An empirical study on challenges for LLM application developers. ACM Transactions on Software Engineering and Methodology, 34(7), 205. https://doi.org/10.1145/3715007

Choung, H., David, P., & Ross, A. (2023). Trust in AI and Its Role in the Acceptance of AI Technologies. International Journal of Human-Computer Interaction, 39(9), 1727-1739. https://doi.org/10.1080/10447318.2022.2050543

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

Garrett, J. J. (2010). The elements of user experience: User-centered design for the Web and beyond (2nd ed.). New Riders Publishing.

Golding, J. M., Lippert, A., Neuschatz, J. S., Salomon, I., & Burke, K. (2024). Generative AI and college students: Use and perceptions. Teaching of Psychology, 52(3), 369–380. https://doi.org/10.1177/00986283241280350

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv. https://doi.org/10.48550/arXiv.2203.05794

Gu, Z., Zhu, Q., He, H., Lan, T., & Yu, Z. (2024). Analysis of emotional cognitive capacity of artificial intelligence. In W. Shen, J. Barthes, J. Luo, T. Qiu, X. Zhou, J. Zhang, H. Zhu, K. Peng, T. Xu, & N. Chen (Eds.), Proceedings of the 2024 27th International Conference on Computer Supported Cooperative Work in Design (CSCWD) (pp. 3182–3187). IEEE. https://doi.org/10.1109/CSCWD61410.2024.10580073

Hadwan, M., Al-Sarem, M., Saeed, F., & Al-Hagery, M. A. (2022). An Improved Sentiment Classification Approach for Measuring User Satisfaction toward Governmental Services' Mobile Apps Using Machine Learning Methods with Feature Engineering and SMOTE Technique. Applied Sciences, 12(11), 5547. https://doi.org/10.3390/app12115547

Kim, J., Klopfer, M., Grohs, J. R., Eldardiry, H., Weichert, J., Cox, L. A., & Pike, D. (2025). Examining faculty and student perceptions of generative AI in university courses. Innovative Higher Education, 50, 1281–1313. https://doi.org/10.1007/s10755-024-09774-w

Li, B., Jiang, G., Li, N., & Song, C. (2024). Research on large-scale structured and unstructured data processing based on large language model. In Proceedings of the International Conference on Machine Learning, Pattern Recognition and Automation Engineering (MLPRAE '24) (pp. 111–116). Association for Computing Machinery. https://doi.org/10.1145/3696687.3696707

Lu, G., Qu, S., & Chen, Y. (2025). Understanding user experience for mobile applications: A systematic literature review. Discover Applied Sciences, 7, 587. https://doi.org/10.1007/s42452-025-07170-3

Nahar, N., Kästner, C., Butler, J., Parnin, C., Zimmermann, T., & Bird, C. (2024). Beyond the comfort zone: Emerging solutions to overcome challenges in integrating LLMs into software products. arXiv. https://doi.org/10.48550/arXiv.2410.12071

Ossai, C. I., & Wickramasinghe, N. (2023). Sentiments prediction and thematic analysis for diabetes mobile apps using Embedded Deep Neural Networks and Latent Dirichlet Allocation. Artificial Intelligence in Medicine, 138, 102509. https://doi.org/10.1016/j.artmed.2023.102509

Law, E. L.-C., Roto, V., Hassenzahl, M., Vermeeren, A. P., & Kort, J. (2009). Understanding, scoping, and defining user experience: A survey approach. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 719–728). https://doi.org/10.1145/1518701.1518813

Schoenmueller, V., Netzer, O., & Stahl, F. (2020). The polarity of online reviews: Prevalence, drivers, and implications. Journal of Marketing Research, 57(5), 853–877. https://doi.org/10.1177/0022243720941832

Shao, Y., Huang, Y., Shen, J., Ma, L., Su, T., & Wan, C. (2025). Are LLMs correctly integrated into software systems? In Proceedings of the 2025 IEEE/ACM 47th International Conference on Software Engineering (ICSE) (pp. 1178–1190). IEEE. https://doi.org/10.1109/ICSE55347.2025.00204

Shata, A., & Hartley, K. (2025). Artificial intelligence and communication technologies in academia: Faculty perceptions and the adoption of generative AI. International Journal of Educational Technology in Higher Education, 22, 14. https://doi.org/10.1186/s41239-025-00511-7

Shen, X., Chen, Z., Backes, M., & Zhang, Y. (2023). In ChatGPT We Trust? Measuring and Characterising the Reliability of ChatGPT. arXiv. https://doi.org/10.48550/ARXIV.2304.08979

Thorne, S. (2024). Understanding the interplay between trust, reliability, and human factors in the age of generative AI. International Journal of Simulation: Systems, Science and Technology, 25(1), 10. https://doi.org/10.5013/IJSSST.a.25.01.10

Wang, J., & Liu, Y. (2023). Deep learning-based social media mining for user experience analysis: A case study of smart home products. Technology in Society, 73, 102220. https://doi.org/10.1016/j.techsoc.2023.102220

Weisz, J. D., He, J., Muller, M., Hoefer, G., Miles, R., & Geyer, W. (2024). Design principles for generative AI applications. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24). Association for Computing Machinery, 378, 1–22. https://doi.org/10.1145/3613904.3642466