ISSN 2616-5775 Vol. 9, Issue 9: 1-7, DOI: 10.25236/AJCIS.2026.090901

A Study on Users' Evaluation of AIGC Quality in Travel Planning

Fukang Fang[1,a], Xue Wang[1,b,*], Guifang Zhou[1]

[1]School of Economics and Management, Hubei University of Automotive Technology, Shiyan, 442002, China
[a]1329313616@qq.com, [b]xuewang_99@163.com
**Corresponding author*

Abstract: *The proliferation of generative AI (GenAI) has led to a growing number of users relying on it for travel planning; however, users' evaluation of AI-generated content (AIGC) quality in travel planning remain unclear. From the perspective of users' subjective perception, this study first draws upon information quality theory, the technology acceptance model and user experience theory to construct an AIGC quality evaluation framework comprising 10 indicators across 3 dimensions: content quality, user experience and itinerary planning. A scenario-based experiment was then employed, in which 4 mainstream Chinese GenAI platforms were selected to complete the same travel planning task, with user-generated content (UGC) from social media serving as a control condition. A questionnaire survey was further conducted to collect users' evaluation of AIGC quality, yielding 211 valid responses. The results show that: first, users' evaluation of AIGC quality exhibit significant group heterogeneity. Specifically, males give significantly lower ratings for personalization, whereas females give significantly lower ratings for comprehensibility and itinerary rationality; users aged 20-30 have higher expectations regarding personalization, whilst those under 20 have higher expectations regarding itinerary rationality, and sensitivity to timeliness increases with age; frequent users rate usefulness higher but assign lower scores to personalization. Second, users' overall evaluation of AIGC quality is significantly better than their evaluation of UGC quality, yet they perceive the timeliness of AIGC as inferior to that of UGC. Third, regression analysis indicates that usage frequency exerts a significant positive influence on users' overall quality evaluation of AIGC, while education level exerts a significant negative influence. This study enriches user-centred research on AIGC quality evaluation and provides user-side empirical evidence for the technological optimisation of AIGC and the intelligent upgrading of tourism information services.*

Keywords: *AIGC, Travel planning, Information quality evaluation, Scenario-based experiment*

1. Introduction

The explosive development of generative artificial intelligence (GenAI), epitomised by ChatGPT, has propelled AI-generated content (AIGC) deep into virtually every domain of human work, life, education, and entertainment. In the domain of travel planning, GenAI can rapidly generate personalized itineraries and attraction recommendations, thereby transforming the traditional way in which travellers plan their trips and saving them substantial time in information search and decision-making. However, both existing research and user feedback indicate that GenAI-assisted travel planning is highly uneven in terms of information accuracy, comprehensiveness, and practicality; inaccurate information may even mislead travel decisions and compromise the user experience[1].

How to evaluate AIGC quality scientifically is one of the keys to advancing AI technology and improving user experience. A review of the extant literature reveals three notable gaps. First, from an evaluation perspective, existing studies have predominantly focused on constructing objective indicator frameworks, such as measures of content correctness and logical coherence. However, as users are the direct recipients of AIGC, their subjective perceptions and evaluation are of critical importance to technological optimization, yet empirical research grounded in users' evaluation remains relatively scarce[2, 3]. Second, in terms of research contexts, most studies have focused on general content scenarios such as news and education. However, there has been little research on quality evaluation specifically targeting the context of travel planning, which involves information integration, temporal and spatial constraints, and personalized needs. Third, in terms of the comparative personalized

ISSN 2616-5775 Vol. 9, Issue 9: 1-7, DOI: 10.25236/AJCIS.2026.090901

dimension, cross-source comparisons between AIGC and user-generated content (UGC) on social media are absent, making it difficult to judge the strengths and weaknesses of AIGC in assisting travel planning.

In view of these gaps, this study takes GenAI-assisted travel planning as its specific context and attempts to answer the following questions: (1) What are the overall characteristics of users' evaluation of AIGC quality in travel planning? (2) What differences exist in AIGC quality evaluation among users with different genders, ages, educational backgrounds, and usage experience, and which factors significantly influence users' overall evaluation? (3) What are the strengths and weaknesses of AIGC compared to UGC?

This study introduces information quality theory, the Technology Acceptance Model, and user experience theory into the AIGC travel-planning context and constructs a subjective quality evaluation index system comprising 10 indicators across three dimensions, thereby enriching user-centred research on AIGC quality evaluation. It also provides evidence from a user perspective for the technological optimisation of AIGC and the intelligent upgrading of tourism enterprises' services.

2. Construction of AIGC quality evaluation index system

2.1. Theoretical foundations

Information quality theory (IQT) emphasises the systematic evaluation of the value of information from multiple dimensions: intrinsic quality (e.g., accuracy and completeness), contextual attributes (e.g., relevance and timeliness), representational characteristics (e.g., readability), and accessibility features (e.g., security)[4, 5]. By providing a multi-dimensional framework for the systematic assessment of information quality, IQT offers a key theoretical basis for evaluating AIGC quality. In addition, in analysing users' subjective evaluation of AIGC quality in travel planning applications, this study draws upon the Technology Acceptance Model (TAM), which foregrounds perceived usefulness and perceived ease of use[6]; user experience theory, which stresses emotional resonance and contextual adaptation beyond functional satisfaction[7]. Together, these theories constitute the theoretical foundation for the construction of the index system and the design of the analytical framework in this study.

2.2. AIGC quality evaluation index system

Building on the above theories, and with reference to the AIGC quality evaluation index system proposed by Guo et al.[8] and the subjective-objective integrated measurement model proposed by Tang et al.[3], this study identifies 10 key evaluation indicators across three dimensions—content quality, user experience, and itinerary planning—in light of the characteristics of the travel-planning scenario, namely information intensity, strong spatio-temporal constraints, and high demand for personalization (see Table 1). The index system takes into account content authenticity, completeness, and personalization from multiple angles, highlights the special requirements of the itinerary-planning context, and reflects users' subjective feelings and satisfaction.

Table 1: Evaluation index system for AIGC quality in travel planning.

Dimension	Indicator	Connotation
Content quality	Credibility	Users' perceived authenticity and reliability of the generated content
	Comprehensiveness	Whether the content covers itinerary, transportation, dining, accommodation, and other aspects
User experience	Personalization	Whether the content fits users' interests, preferences, and needs
	Comprehensibility	Whether the content is clear, concise, and easy to understand
	Human-likeness	Whether the generated content exhibits a natural language style close to human expression
	Usefulness	The extent to which the content actually helps users make travel plans
	Inspiration	Whether the content stimulates users' inspiration or provides new ideas
Itinerary planning	Timeliness	The update frequency and real-time nature of the information (opening hours, ticket prices, transportation, etc.)
	Itinerary rationality	The logical coherence of the planned itinerary and the rationality of time allocation
	Itinerary attractiveness	Whether the itinerary includes attractions or activities of interest to users

ISSN 2616-5775 Vol. 9, Issue 9: 1-7, DOI: 10.25236/AJCIS.2026.090901

3. Research design

3.1. Questionnaire design

A questionnaire survey was adopted to collect users' evaluation of AIGC quality in travel planning. The questionnaire consisted of five parts: (1) basic information (gender, age, education, and occupation); (2) AIGC usage experience (whether respondents had used AIGC for travel planning, usage frequency, and commonly used platforms); (3) scenario-based experiment ratings: each respondent was randomly assigned to view one of five scenarios and rate 10 evaluation indicators on a five-point Likert scale; the random assignment was designed to avoid fatigue from repeated evaluation (see Section 3.2); (4) problem feedback (problems encountered in using AIGC and the aspects most in need of improvement); and (5) comparative evaluation and continuance intention (quality comparison between AIGC and UGC, perceived advantages of AIGC over self-planning, and intention to continue using AIGC).. All scale items were adapted from mature instruments in prior research and revised through a pilot survey.

3.2. Scenario-based experiment design

To ensure the comparability of the evaluation objects, a scenario-based experiment was employed. Based on monthly active user rankings and social attention, four mainstream Chinese GenAI platforms—Doubao, Kimi, ERNIE Bot, and DeepSeek—were selected. An identical prompt was submitted to each platform: This weekend, my best friend and I are going to spend a day in Wuhan; please help me create a travel itinerary. The four resulting travel plans constituted four AIGC scenarios. In addition, an original, high-quality travel planning guide posted by an influential user on Xiaohongshu, a major Chinese social media platform, was selected as Scenario 5, forming the UGC control condition for comparing quality differences between AIGC and UGC.

3.3. Data collection and sample characteristics

A pilot survey was conducted before the formal survey to examine the appropriateness of the questionnaire structure and item wording. The formal survey was distributed through social media and travel forums, yielding 239 responses in total. After excluding 28 invalid responses in which all items were answered identically or in which responses were obviously contradictory, 211 valid samples were obtained, for an effective rate of 88.3%. The sample characteristics were as follows: females accounted for 61% and males for 39%; 74% of respondents were aged 20-30; 69% held a bachelor's degree, while high school education or below, junior college, and master's degree or above accounted for 12%, 10%, and 9%, respectively; students accounted for 80%; 87% of respondents had used AIGC for travel planning, of whom 44% were multiple-time users. The sample was dominated by young, highly educated respondents, which is consistent with the overall profile of GenAI users and can therefore reflect, with reasonable fidelity, the subjective perceptions of target users regarding the AIGC quality in travel planning.

4. Empirical results and analysis

4.1. Reliability and validity tests

In terms of reliability, the Cronbach's alpha coefficients of the five scenario scales (Doubao, Kimi, ERNIE Bot, DeepSeek, and Xiaohongshu UGC) were 0.888, 0.862, 0.848, 0.767, and 0.749, respectively; the alpha coefficient of the merged sample scale was 0.859, and that of the usage-problem scale was 0.864—all above the acceptable threshold of 0.7, indicating good internal consistency. In terms of validity, the Kaiser-Meyer-Olkin (KMO) value of the merged sample was 0.840, and Bartlett's test of sphericity was significant ($p < 0.001$), indicating that the data were highly suitable for factor analysis. The KMO values of the individual platform scales ranged from 0.626 to 0.721 (0.866 for the usage-problem scale), all reaching acceptable levels. The questionnaire items were revised on the basis of validated scales and in combination with the pilot survey and expert discussions, ensuring content validity.

ISSN 2616-5775 Vol. 9, Issue 9: 1-7, DOI: 10.25236/AJCIS.2026.090901

4.2. Comparison of AIGC and UGC quality evaluations

The four AIGC scenarios were merged and compared with the UGC control condition (see Table 2). Of the 10 indicators, AIGC received higher ratings than UGC on 9. The advantages were most pronounced for personalization (4.17 vs. 3.63), human-likeness (4.23 vs. 3.47), inspiration (4.01 vs. 3.21), and itinerary rationality (3.86 vs. 2.88). The overall quality mean of AIGC (3.92) was higher than that of UGC (3.44). Only in timeliness did UGC (3.79) outperform AIGC (3.49), reflecting UGC's inherent advantage in information freshness, which derives real-time sharing by real users. Overall, users' evaluation of AIGC quality surpassed their evaluation of UGC quality, yet timeliness remains an area in which AIGC needs improvement.

Table 2: Comparison of perceived scores of AIGC and UGC across indicators.

Indicator	AIGC mean	UGC mean	Difference
Credibility	4.01	3.56	+0.45
Comprehensiveness	4.16	3.44	+0.72
Personalization	4.17	3.63	+0.54
Comprehensibility	3.95	3.53	+0.42
Human-likeness	4.23	3.47	+0.76
Usefulness	3.95	3.49	+0.46
Inspiration	4.01	3.21	+0.80
Timeliness	3.49	3.79	−0.30
Itinerary rationality	3.86	2.88	+0.98
Itinerary attractiveness	4.05	3.51	+0.54
Overall quality	3.92	3.44	+0.48

4.3. Group difference tests

Independent-samples t tests were conducted to examine differences in users' scores on the 10 indicators between genders, and one-way ANOVAs were used to examine differences across age, education, and usage frequency groups. The results are summarized in Table 3. Overall, the quality evaluation of different user groups display a clear pattern of stratified demands.

Table 3: Summary of group difference tests in AIGC quality evaluation.

Grouping variable	Indicators with significant differences ($p < 0.05$)	Direction and characteristics of differences
Gender	Personalization; comprehensibility; itinerary rationality	Females scored personalization significantly higher (4.21 vs. 3.31); males scored comprehensibility (4.52 vs. 3.73) and itinerary rationality (3.54 vs. 3.26) significantly higher
Age	Personalization; timeliness; itinerary rationality	Among personalization scores, users aged 20-30 scored lowest (3.76); timeliness scores decreased with age (4.63 → 2.44); among itinerary rationality scores, users under 20 scored lowest (2.78)
Education	Credibility; comprehensibility; human-likeness; timeliness	The high-school-and-below group scored highest on credibility (4.38) and comprehensibility (4.29); the bachelor's group scored highest on human-likeness (4.39); the junior-college group scored highest on timeliness (4.15)
Usage frequency	Overall quality; usefulness; itinerary rationality; comprehensibility; personalization	Scores on overall quality, usefulness, itinerary rationality, and comprehensibility increased significantly with usage frequency; personalization showed a reverse gradient, with the infrequent-use group (4.07) scoring higher than the frequent-use group (3.66)

In terms of gender, indicators such as credibility, comprehensiveness, and human-likeness shows no significant differences, suggesting that the baseline perceptions of AIGC quality are stable across genders. However, females respondents gave significantly lower ratings than males on comprehensibility and itinerary rationality, indicating that females have higher expectations for concise expression and logically coherent itineraries, whereas males gave significantly lower ratings to personalization, suggesting that males seek more personalized arrangements rather than template-based experiences such as social-media-famous check-in spots.

In terms of age, users aged 20-30 expected greater personalization from AIGC; sensitivity to

ISSN 2616-5775 Vol. 9, Issue 9: 1-7, DOI: 10.25236/AJCIS.2026.090901

timeliness increased with age; and users under 20 had stricter requirements for itinerary rationality. These findings reflect differences in the perceptual focus on AIGC quality across age groups.

In terms of education, respondents with a bachelor's degree or above gave lower ratings to and applied stricter standards to credibility and comprehensibility, consistent with their richer knowledge base and stronger information discernment; the bachelor's-degree group's high requirements for human-likeness reflect their expectations of emotionally resonant expression.

In terms of usage frequency, frequent users rated the utilitarian dimensions of AIGC significantly higher, corroborating the cumulative effect of usage experience on perceived quality. However, frequent users are more sensitive to template-based content, and their personalization scores were in fact lower than those of infrequent users. The tension between personalization and template-based output is a key challenge in retaining experienced users.

4.4. Correlation and regression analysis of influencing factors

Correlation analysis showed that usage frequency was significantly and positively correlated with users' overall evaluation of AIGC quality ($r = 0.328$, $p < 0.01$), while education was weakly and negatively correlated with the overall evaluation ($r = -0.149$, $p < 0.01$). Correlations between gender ($r = 0.039$), age ($r = -0.010$), and occupation ($r = 0.089$) and the overall evaluation were all non-significant. Furthermore, multiple linear regression analysis was performed with gender, age, education, occupation, and usage frequency as independent variables and users' overall evaluation of AIGC quality as the dependent variable (see Table 4). The results showed that the model was significant overall. Usage frequency exerted a significant positive influence on the overall evaluation ($B = 0.280$, $\beta = 0.349$, $p < 0.001$), constituting the dominant factor in users' perceived evaluation; education exerted a significant negative influence ($B = -0.120$, $\beta = -0.186$, $p < 0.05$), indicating that the higher the education level, the stricter the evaluation criteria. Gender, age, and occupation were not significant. These results suggest that, compared with stable demographic attributes, usage frequency and education are the key factors influencing users' overall quality evaluation of AIGC.

Table 4: Regression results of user characteristics on the overall evaluation of AIGC quality (N = 211).

Variable	B	SE	β	t	p	VIF
Constant	3.839	0.267	—	14.404	<0.001	—
Gender	0.012	0.083	0.012	0.149	0.882	1.031
Age	−0.106	0.081	−0.127	−1.302	0.195	1.595
Education	−0.120	0.054	−0.186	−2.229	0.027	1.177
Occupation	0.151	0.130	0.118	1.165	0.246	1.739
Usage frequency	0.280	0.064	0.349	4.379	<0.001	1.066

5. Discussion

First, the usage experience play a dominant role. The regression results show that usage frequency is the only strongly significant positive factor influencing the overall quality evaluation, which is consistent with TAM's core proposition that usage experience shapes technology perception. After becoming proficient in functions such as generating alternative plans and dynamically adjusting itineraries, frequent users can fully leverage AIGC's information-integration value and therefore form higher-quality evaluations. This finding implies that cultivating users' usage habits and designing onboarding support for novices are as important as improving the technology itself.

Second, users' evaluation of AIGC quality show group differentiation. Female users' stronger preference for comprehensibility and itinerary rationality, and male users' focus on personalization, can be understood through the differences between emotional design and functional appeals in user experience theory. The demand for itinerary rationality among users under 20, for personalization among users aged 20-30, and for timeliness among users over 40 reflects differences in digital lifestyles and travel preferences across age groups. The negative education effect corroborates IQT's proposition that evaluation criteria tighten as knowledge levels rise. The decline in personalization scores among frequent users reveals an "expectation paradox" in AIGC personalization services: experienced users, having been exposed to a larger number of generated samples, are less tolerant of template-based outputs. Taken together, these findings indicate that AIGC quality evaluation exhibits significant group heterogeneity, and that a "one-size-fits-all" output strategy cannot meet the needs of different market segments.

ISSN 2616-5775 Vol. 9, Issue 9: 1-7, DOI: 10.25236/AJCIS.2026.090901

Third, AIGC and UGC play complementary rather than substitutive roles in information quality positioning. AIGC's substantial lead in structured-integration dimensions such as comprehensiveness, itinerary rationality, and inspiration stems from its large-scale data processing and logical organization capabilities, whereas UGC's advantage in timeliness stems from real-time experiential sharing by real users. This suggests that the optimal form of tourism information services is not to replace UGC with AIGC but to build a hybrid ecosystem combining an AIGC integration framework with UGC real-time feedback.

6. Implications

First, implement segmented service strategies for different user groups. Platforms can refine their output strategies based on user profiling: for male users, strengthen interest-tag-driven personalized recommendations and experiential activities; for female users and working professionals aged 30-40, optimize conciseness of expression and handling of time constraints, and provide precise time-allocation and congestion-avoidance advice; for frequent users, provide custom templates and advanced parameter-tuning options to alleviate template fatigue; and for highly educated users, improve the logical rigor and source traceability of content.

Second, technology developers should prioritize addressing shortcomings in accuracy and timeliness. In response to user feedback on information accuracy, itinerary rationality, and timeliness, the risk of hallucination can be reduced by introducing real-time data interfaces (transportation, ticketing, opening hours), retrieval-augmented generation (RAG) technology, and source-attribution mechanisms. Natural language processing and machine learning methods should be used to deepen user profiling, and dynamic recommendation mechanisms combining scenario features with emotional factors should be designed to raise the level of personalization while reducing templating. A user-feedback loop should be established to incorporate behavioral data such as itinerary-modification records into model iteration.

Third, tourism enterprises should build an "AIGC + UGC" integrated service ecosystem. AIGC's data-analytics capabilities can be used to mine user behavior preferences and form themed planning templates and personalized service modules; authentic tourist reviews and experiential content can be embedded into the AIGC generation process, so that "intelligent recommendation + authentic feedback" enhances the accuracy and timeliness of information; and self-service custom templates can be provided for frequent travelers, thereby covering the needs of diversified market segments.

Fourth, regulators should improve AIGC quality certification and data-governance standards. Regulators should draw on relevant industry white papers and standards to formulate scientifically grounded quality-evaluation criteria and establish a quality-certification and labeling system for AIGC tourism content. Platform data sharing and service-quality disclosure should be promoted, and user privacy and data security should be effectively safeguarded while services are optimized. Through collaborative governance among developers, tourism enterprises, and regulators, AIGC quality and the intelligent upgrading of tourism information services can be continuously improved.

7. Conclusions

This study developed an evaluation framework of AIGC quality in the context of AI-assisted travel planning, and utilised scenario-based experimental method to investigate and analyse users' evaluation of AIGC quality. This study enriches user-centred research on AIGC quality evaluation and provides user-side empirical evidence for the technological optimisation of AIGC and the intelligent upgrading of tourism information services. However, several limitations should be acknowledged. The sample was dominated by young people aged 20-30 and by students, with insufficient coverage of middle-aged and older users and users from other cultural backgrounds. The experimental scenario was limited to a one-day trip to Wuhan, and more complex planning scenarios—such as long-distance multi-city trips, family trips, and business trips—were not examined. The evaluation system focused on subjective perceptions and did not incorporate emotional experience. Although the scenario-based experiment enabled a controlled comparison, the single-task design limits external validity. Future research should explore these dimensions further, promoting the evolution of tourism information services toward greater intelligence and human-centeredness.

ISSN 2616-5775 Vol. 9, Issue 9: 1-7, DOI: 10.25236/AJCIS.2026.090901

References

[1] Fu, C. J., Silalahi, A. D. K., Shih, I. T., et al. (2024). Assessing ChatGPT's information quality through the lens of user information satisfaction and information quality theory in higher education: A theoretical framework. Human Behavior and Emerging Technologies, 1, 8114315.

[2] Song, S. J., Zhao, Y. X., & Zhu, Q. H. (2023). From ELIZA to ChatGPT: AI-Generated Content(AIGC) Credibility Evaluation in Human-Intelligent Interactive Experience. Information and Documentation Services, 44(4), 35-42.

[3] Tang, K., Li, B. Y., & Zhang, X. Y. (2024). Cross-modal Artificial Intelligence Generative Content Quality and Efficacy Measurement Based on Subjective-Objective Convergence. Information Studies: Theory & Application, 47(11), 150-161.

[4] Sun X. N., Zhang, S, & Guo W. Z. (2025). A comparative study of perceived information value between knowledge-enhanced large language models and search engine users: based on task complexity and information acquisition scenarios. Library and Information Service, 69(12):90-103.

[5] Sun X. N., & Zhang, S. (2026). The Construction of Answer Quality Evaluation Index System for Conversational Search System: Based on DeepSeek. Information Studies: Theory & Application, 49(2), 168-178.

[6] Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319-340.

[7] Hassenzahl, M., & Tractinsky, N. (2006). User experience-A research agenda. Behaviour & Information Technology, 25(2), 91-97.

[8] Guo, Y. J., Liu, Z. Y., Feng, S. Q., et al. (2024). Research on the Construction of Quality Evaluation Index System of AIGC——Taking ChatGPT as an Example. Library and Information Service, 68(22), 93-104.