Beyond the View: A Multi-Source Framework for Tourism Risk Identification, Reverse Recommendation, and AIGC Decision Content

Bowen Lei, Jingyi Xie, Wanting Qiu, Zijie Wang, Yunran Xing, Yunuo Luo, Sijin Song, Liang Hong*

College of Media Engineering, Communication University of Zhejiang, Hangzhou, China

*Corresponding Author

Abstract: Online platforms and generative artificial intelligence have changed how tourists collect, compare, and interpret travel information. User-generated content (UGC) contains concrete descriptions of queues, prices, service encounters, environmental conditions, and other situations that may matter before a trip. Yet tourism information services still tend to organize information around attraction introduction, review aggregation, and positive recommendation. The decision problem addressed here is different: how can scattered negative or cautionary experiences be converted into usable risk information without turning them into a simple judgment of destination quality? This paper proposes a risk-oriented framework that links multi-source UGC, large language models (LLMs), reverse recommendation, and AIGC decision content. The framework defines tourism risk information through five core elements—tourism object, scenario, risk type, risk manifestation, and decision prompt—and uses the five-element record as the bridge between narrative reviews and later decision processing. LLM-based semantic analysis is used to identify risk-related expressions, while rule constraints and evidence fields are introduced to limit unsupported interpretation. The recommendation stage starts from risks a user wants to avoid, filters options that conflict with those constraints, and then searches for alternatives using tourism-attribute similarity, user-need similarity, and risk-feature difference. AIGC is placed after the analytical stages so that generated content is grounded in structured evidence rather than treated as a factual source by itself. The study is conceptual and methodological. It does not report completed deployment, benchmark accuracy, or user-effect measurements. Its boundary is deliberately set at the design of an interpreted information-processing chain and the definitions needed for later empirical validation.

Keywords: Multi-Source Data Fusion; User-Generated Content; Large Language Models; Tourism Risk Identification; Reverse Recommendation; AIGC; Decision Support

1. Introduction

Digital platforms have changed how tourists search for information and form expectations before travel. Reviews and social media posts can contain details that official descriptions may not capture, such as holiday queues, unexpected charges, service delays, or environmental conditions. Tourism research therefore treats online reviews as an important form of electronic word-of-mouth [1,2]. The 14th Five-Year Plan for Tourism Development also calls for smart tourism and wider use of digital technologies [3]. Research on crowded scenic areas also links risk perception with tourist behavior and safety [4]. Aspect-based analysis also shows the value of separating different aspects within tourist reviews [1].

The problem is not simply that tourists lack information. They may have too much of it, while the most decision-relevant details are scattered across different sources. A conventional recommendation interface tends to ask what a person likes and then ranks attractive options. A risk-sensitive tourist may ask a different question: "What should I avoid under my conditions?" The answer may depend on time, crowd level, price sensitivity, service expectations, or other situational constraints. Negative online information can have diagnostic

Copyright @ STEMM Institute Press

value in such decisions, and research on crowded scenic areas also shows the relevance of risk perception to tourist behavior and safety [4,5]. UNWTO work on visitor concentration further shows why time-based and spatial management can matter when tourism pressure becomes concentrated [6].

Large language models provide a practical way to process the linguistic variation found in UGC. A single risk may be expressed through many forms: "waited nearly two hours," "the line was endless," or "do not come here during the holiday afternoon." Keyword matching can capture some of these phrases, but it does not by itself establish the object, situation, manifestation, and possible decision implication. Recent work on tourism UGC, generative AI, and information adoption suggests that AI can support richer information processing, while also raising questions about evidence quality and model reliability [7-10]. The research problem addressed in this paper is therefore the missing middle layer between raw tourism narratives and user-facing decision support.

The framework is not a claim that an LLM can automatically discover objective tourism risks. It is a design for organizing evidence. Its core unit is a structured risk record that links the tourism object and situation to a risk type, observable manifestation, and decision prompt. The same representation can then be passed to recommendation and content generation.

The paper has three closely connected contributions. It defines a five-element representation for tourism risk information; it develops a reverse recommendation logic in which risk constraints are used before alternative matching; and it positions AIGC as a controlled communication layer that converts structured evidence into readable decision content. The study remains within a methodological boundary. No completed system deployment, accuracy result, causal effect, or user-satisfaction improvement is claimed here.

2. Related Research

2.1 Tourism UGC and Online Review Research

Online reviews preserve experience descriptions in relatively natural language and have been used to study tourism decisions, information evaluation, destination image, and satisfaction [1,2]. Tourism text can also reveal recurring experience dimensions and positive or negative perceptions [11,12]. These studies support UGC as an information source, but they do not by themselves extract decision-relevant risk conditions. Sentiment analysis of tourism comments also shows how review data can support destination analysis [12]. Destination-image analysis based on online text provides a related example [11].

UGC is heterogeneous across platforms, users, and writing styles. A multi-source design can broaden coverage, but source differences should remain visible. The framework therefore keeps source and context in the evidence chain. Repeated statements may increase attention to a condition, but repetition alone does not establish permanence or causality [2].

2.2 Online Review Usefulness, Structure, and Risk Information

Tourists can face information overload when reading large numbers of reviews. Review quality, diagnostic value, and perceived usefulness affect whether such information can be used effectively [2,13]. Structured reviews can also improve information depth and diagnostic information for experience-oriented products [13]. This supports the present approach: narrative information should be converted into a form in which object, context, risk condition, and possible response can be compared.

Negative information deserves attention because it can reveal conditions under which an experience may not meet a visitor's expectations. Wang and Wang show that negative online interaction information can influence heterogeneous tourism consumers through perceived risk [5]. The present study takes a narrower position. Negative sentiment is a signal, not a definition of risk. "I did not enjoy the atmosphere" may describe a personal preference, whereas "we waited about two hours to enter during the holiday" contains a concrete condition that can be useful for another tourist. The distinction is central to the proposed risk representation.

2.3 Large language Models and AIGC in Tourism

Generative AI is increasingly discussed in tourism information services, content production, and consumer decision processes. Studies have examined tourism information

 Copyright @ STEMM Institute Press

adoption [7], industry applications and risks [8], AIGC-based content production [14], domain-specific UGC entity recognition [9], and LLM-supported tourist value co-creation [10]. Industry analysis also describes AI-supported travel planning and personalized decision support as emerging applications [15].

The same technologies introduce a boundary that matters for this study. An LLM can make a sentence clearer while also adding a detail that was never present in the input. For a tourism decision system, that is not a minor stylistic error: an invented price, waiting time, opening hour, or safety condition can change a decision. The proposed framework therefore assigns AIGC a downstream role. Structured evidence comes first; generation follows; verification remains necessary. Systematic review work also identifies consumer-level benefits and risks of generative AI in hospitality and tourism [16].

2.4 Research Gap

The literature provides foundations for review analysis, structured information, risk perception, and generative AI, but the complete path between them is less explicit. A decision-support system needs more than sentiment or destination image: it needs the condition, scenario, and possible decision implication. Conventional recommendation emphasizes preference matching, while risk avoidance is less often treated as a direct filtering objective. This framework makes the transformation between these stages explicit.

3. Research Framework

3.1 Overall Framework

Figure 1 presents the proposed research framework. It is organized as an information transformation chain rather than as a list of independent functions. Multi-source UGC is first standardized so that later semantic processing has a consistent input. Risk-related semantics are then extracted with the help of an LLM and constrained output rules. The extracted information is represented as a five-element risk record. That record becomes a decision constraint for risk filtering and alternative matching. AIGC is applied only after the analytical information has been structured and checked, producing content that can be read by tourists or used in tourism communication.

A feedback path is also included for later validation. User feedback, error cases, and verification results could refine category definitions, prompts, and filtering rules. This is a planned mechanism, not an achieved learning result in the current study.

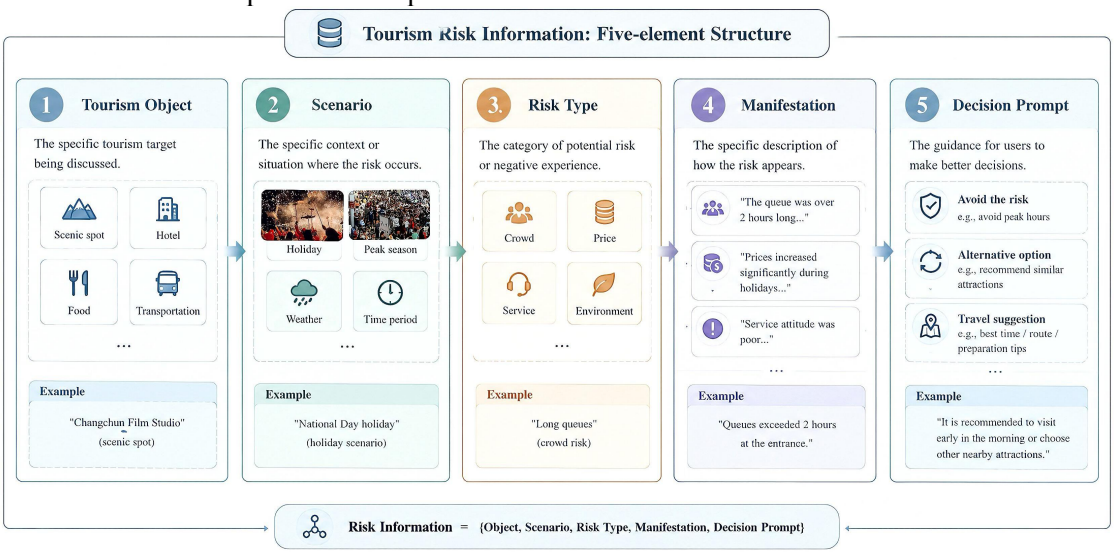

Figure 1. Overall Research Framework

3.2 Research Boundary and Design Principles

The research boundary is deliberately narrow. The framework focuses on review-supported conditions, especially crowding and queuing, price and hidden costs, service experience, and environment or overall experience. It does not predict accidents, assess legal status, or determine objective destination quality. A single negative review is not treated as proof of permanent risk. The unit of analysis is a decision-relevant risk statement in context.

Three principles follow from this boundary. Risk identification should remain evidence-oriented, with a supporting phrase or record for each

claim. Risk filtering should be conditional: an option may conflict with a user's needs without being "bad" in general. Generated content should remain traceable to structured input. "AI" is therefore not used as a substitute for missing evidence.

4. Tourism Risk Identification

4.1 Risk Classification

Based on the decision situations emphasized in the project design, four primary categories are used here: crowding and queuing, price and hidden-cost concerns, service experience, and environment and overall experience. They are not an exhaustive taxonomy. They are a practical starting point for organizing conditions that can influence time cost, economic cost, service expectations, and experiential suitability. The categories can later be expanded if empirical data reveal stable risk patterns outside the current scope.

4.2 Risk Semantic Identification

As is shown in Table 1, the main input is a review or a factual description extracted from a review. The LLM is asked to decide whether the text contains a risk-related statement, identify its category, describe the observable manifestation, and produce a decision prompt. An evidence phrase is retained alongside the core record. It is not one of the five core elements; its role is traceability. This distinction matters because a generated summary should not become detached from the original observation.

Table 1. Proposed Tourism Risk Categories

Risk category	Typical manifestations and decision relevance
Crowding and queuing	Long queues, high density, congestion, and peak-period pressure; relevant to time cost and visit comfort.
Price and hidden cost	Price increase, poor value, additional or unexpected charges; relevant to economic cost and budget uncertainty.
Service experience	Service attitude, response delay, and operational instability; relevant to service expectation and reliability.
Environment and experience	Cleanliness, noise, facilities, and environmental conditions; relevant to experience quality and suitability.

Risk should not be inferred from negative sentiment alone. Consider two statements: "I personally prefer quieter places, so this attraction felt too lively" and "we waited about two hours before entering during the holiday." The first describes an individual preference and does not by itself establish a crowding condition. The second contains a time-bound, observable manifestation and can be represented as a potential crowding risk. The model is therefore expected to separate personal evaluation from decision-relevant conditions.

The extraction task maps narrative evidence to fixed fields. Entity recognition can identify destinations or services, while semantic analysis links the statement to a risk category and manifestation. The LLM handles linguistic variation; category definitions and the output schema provide the decision structure for later checking.

4.3 Five-Element Representation of Tourism Risk Information

The central concept of this paper is the five-element representation: tourism object, scenario, risk type, risk manifestation, and decision prompt. Together they form a compact description of a risk statement that can be passed from semantic analysis to recommendation and content generation. In formal terms, Risk Information = {Object, Scenario, Risk Type, Manifestation, Decision Prompt}.

Tourism object answers what or whom the information refers to: a scenic spot, hotel, restaurant, route, or other tourism service. Scenario records when or under what conditions the issue appears, such as a holiday afternoon, weekend, peak season, or a particular service process. Risk type provides the category used for filtering, such as crowding or price. Risk manifestation records how the condition appears in observable terms; quantitative details such as "two-hour queue" or "additional charge" should be preserved when the source provides them. Decision prompt is different from a risk label. It translates the information into a possible decision concern, for example "check peak-period conditions" or "compare alternatives before booking."

A tourism risk is rarely meaningful without context. "Long queue" is incomplete without the object, scenario, or decision implication. The record also separates evidence from evaluation: a user can see that a condition was reported without being told that the destination is universally poor. Figure 2 shows this movement from experience to decision-oriented information.

 Copyright @ STEMM Institute Press

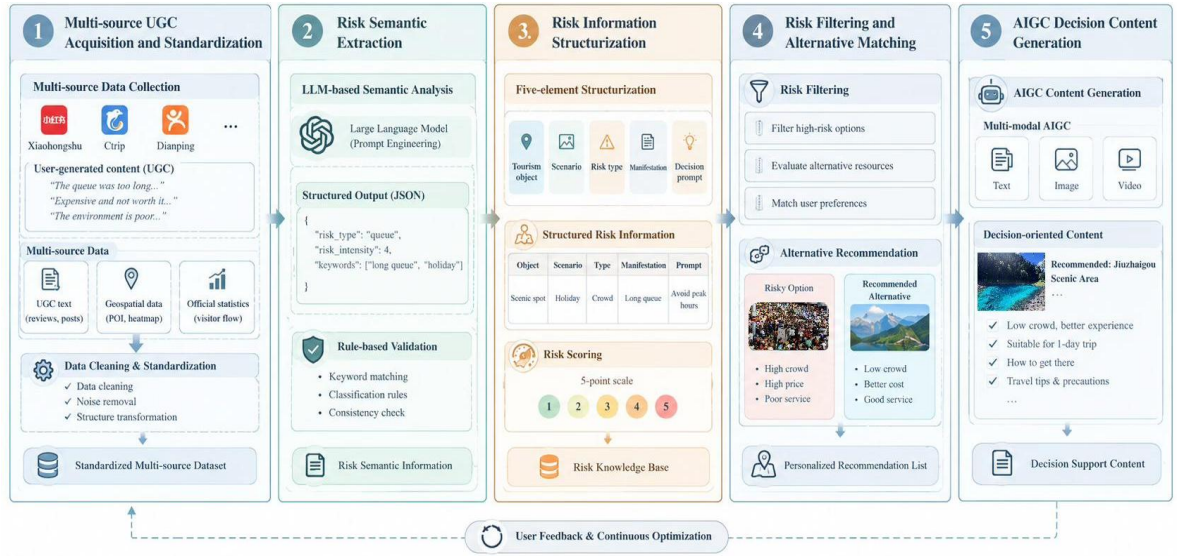

Figure 2. The Five-Element Structure of Tourism Risk Information

5. Risk-Oriented Reverse Recommendation

5.1 From Positive Recommendation to Risk Filtering

As is shown in Figure 3, traditional recommendation often begins with preference: a user states an interest, candidate destinations are collected, and the system ranks them. The reverse recommendation mechanism proposed here changes the starting point. It asks which conditions the user wants to avoid and treats those conditions as constraints before ranking alternatives. In simplified form, the process is risk identification → risk filtering → alternative matching. This is not a claim that risk filtering produces a universally safer destination. It is a change in the information-processing order: potentially unsuitable options are removed or deprioritized before the system spends effort comparing alternatives.

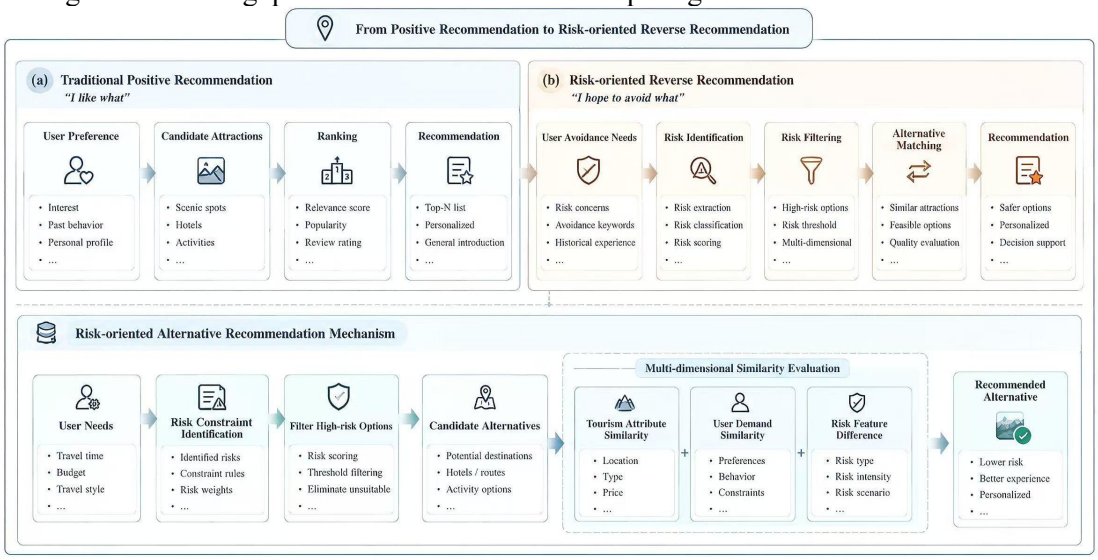

Figure 3. Risk-Oriented Reverse Recommendation and Alternative Matching Mechanism

5.2 Risk Quantification

Based on the extracted risk information, this study represents risk intensity on a five-point scale. For each tourism object, the average score of each risk category is calculated from the identified UGC and aggregated by category weights to obtain an overall risk value:

$$R = \sum_{k=1}^{K} w_k r_k \quad (1)$$

where R denotes the overall tourism risk value; K is the number of risk categories; r_k is the average score of category k; and w_k is its weight, with the weights summing to one. The project uses four categories: crowding and queuing, price and hidden costs, service experience, and environment and overall experience. A higher five-point score indicates stronger risk-related evidence in UGC, not an objective probability of harm or a universal quality judgment.

At the present stage, the category weights are adjustable parameters rather than empirically optimized values. The risk value formalizes the project design but is not a validated predictive model. Future work can revise the weights using an annotated benchmark and user feedback.

5.3 Alternative Matching Mechanism

Filtering alone does not solve the decision problem. A tourist seeking cultural sightseeing but wishing to avoid holiday queues still needs alternatives. Matching therefore considers tourism-attribute similarity, user-need similarity, and risk-feature difference, linking the desired experience with the risk that triggered filtering.

The conceptual matching function is expressed as:

$$M = \alpha Sa + \beta Sn + \gamma Dr \quad (2)$$

where M denotes the alternative matching score; Sa denotes tourism-attribute similarity; Sn denotes user-need similarity; Dr denotes the difference in relevant risk features; and α, β, and γ are adjustable weights. Equation (2) describes the matching logic rather than a calibrated prediction model. The weights have not been estimated from a verified benchmark database, so the score is not presented as validated recommendation accuracy.

5.4 Risk Information is not a Value Judgment

This distinction is essential to the meaning of "risk" in the framework. A crowded attraction may still have high cultural value; a higher price may accompany a distinctive experience; and a service complaint may be isolated. The framework therefore reports risk as a condition attached to a scenario, not as an overall destination score. A reverse recommendation can say that an option conflicts with a user's stated avoidance need while leaving the broader value of that option open. This makes the result more transparent and reduces the tendency to turn negative UGC into an automatic destination ranking.

6. AIGC-Assisted Decision Content Generation

6.1 From Structured Information to Decision Content

The second project line is incorporated at the final information-service layer. Its role is not to duplicate risk identification. Instead, it takes structured risk information and recommendation evidence and converts them into readable decision content. This separation is important because generation and analysis serve different purposes: the analytical layers organize evidence and constraints, while AIGC explains those results in a form that a tourist can use.

6.2 Generation Process

Figure 4 shows the proposed content-generation process. Structured risk records and candidate alternatives are assembled as the input. A prompt template then specifies the task, output form, evidence requirements, and restrictions on unsupported claims. The LLM generates a draft such as a risk reminder, comparison note, alternative recommendation, or short tourism article. The draft is checked against the structured evidence and applicable rules before it is released. If a claim cannot be traced to the available information, it should be removed, qualified, or marked for verification.

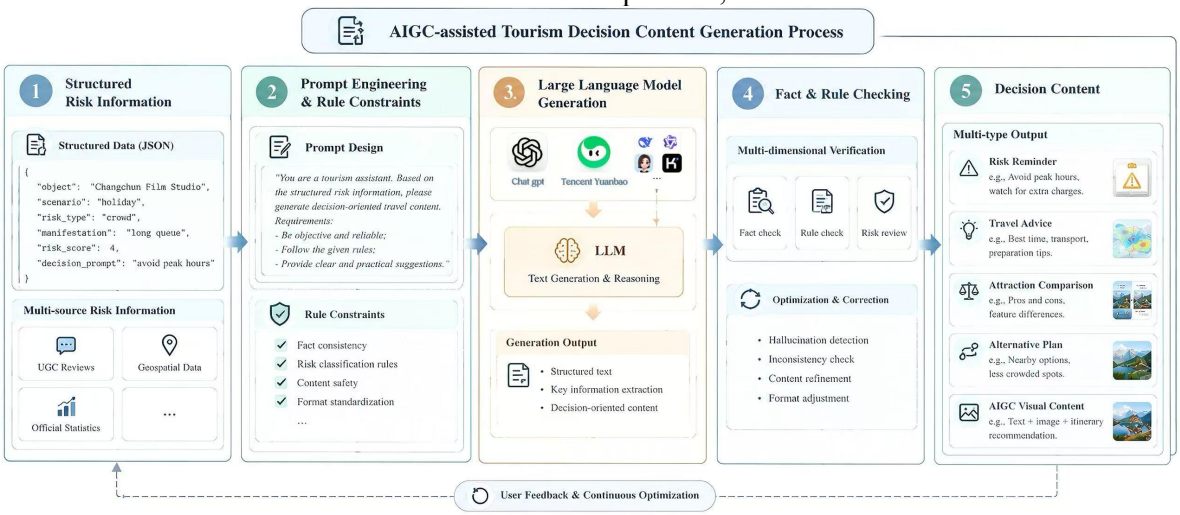

Figure 4. AIGC-Assisted Tourism Decision Content Generation

 Copyright @ STEMM Institute Press

6.3 Content Forms

The output can take several forms without changing the underlying analytical chain: "what to watch for before visiting," peak-period reminders, comparisons between similar attractions, explanations of why an alternative may fit a stated need, or AIGC-based tourism graphic and article content. For county-level tourism scenarios, the same structure can organize scattered local evaluations into decision-oriented material. These are proposed output forms, not evidence that a completed content-production system has already been deployed in this study.

6.4 AIGC Boundaries

AIGC should not fill gaps in evidence. Prices, waiting times, opening hours, transport conditions, and other time-sensitive facts should come from verified inputs rather than model memory. The model can organize and explain supplied information. When evidence is insufficient, the output should be a verification notice rather than an invention. This boundary keeps content generation subordinate to the evidence structure [8,16].

7. Discussion

7.1 Subjectivity and Platform Heterogeneity

UGC reflects individual experience and platform populations. Repetition across sources may deserve attention, but it does not prove permanence, causality, or representative elections Source metadata should therefore be preserved where possible. Multi-source fusion means combining evidence while retaining context [2].

7.2 Risk versus Sentiment

Sentiment can help locate potentially important passages, but it is insufficient as the final risk definition. The five-element representation requires the system to identify an object, scenario, category, manifestation, and decision prompt. This gives a more concrete basis for later filtering than a positive or negative label alone. The approach is compatible with research showing that tourism reviews contain multiple perception dimensions and that negative information can have decision relevance [5,11].

7.3 LLM Uncertainty and Rule Constraints

LLMs make semantic processing flexible, but their outputs may vary and summaries may contain unsupported details. The response is to constrain what the model can output. Fixed categories, the five-element schema, evidence phrases, prompt rules, and secondary checks provide several points for error detection. The lightweight design can later be compared with supervised models or annotated benchmarks.

7.4 Relationship Between the two Project Lines

The two project lines are connected as successive parts of one information process. The first provides the risk-oriented decision logic: multi-source information is transformed into risk records, filtered through user constraints, and used to identify alternatives. The second contributes the AIGC content layer, where structured analytical outputs become user-facing decision material. The combined chain can be summarized as UGC → risk information → decision support → AIGC content. This relationship keeps the two research directions distinct while giving them a common information interface.

7.5 Limitations and Future Validation

Several boundaries remain. The four risk categories and five-element representation need testing on larger datasets, while risk intensity and matching weights remain uncalibrated. Risk changes over time, platform sampling, and source reliability also require study. The framework does not claim that risk-oriented recommendation improves satisfaction, safety, or decision quality; those outcomes require controlled evaluation. Future work should build a verified benchmark, test extraction consistency, compare constrained generation methods, and evaluate the usefulness and traceability of structured risk content.

8. Conclusion

This paper proposed a risk-oriented framework that connects multi-source tourism UGC, LLM-based semantic processing, structured risk information, reverse recommendation, and AIGC decision content. Its central idea is to make potential problems visible as contextual information rather than treating negative reviews as a direct destination judgment. The five-element representation provides the common language across the stages: it records what the

Copyright @ STEMM Institute Press

risk refers to, under what scenario it appears, what type of risk is involved, how the risk manifests, and what decision prompt follows from the evidence.

The framework also changes the role of recommendation. Instead of moving directly from preference to ranking, it allows users to state avoidance needs and uses those needs to filter options before matching alternatives. AIGC is placed after this analytical process, so generated content can explain structured evidence instead of acting as an unrestricted factual source. The result is a lightweight methodological path rather than a completed empirical system. Its value will ultimately depend on later validation with real multi-source data, annotated risk records, calibrated matching parameters, and user studies.

Acknowledgments

This paper is supported by 2026 National College Students' Innovation and Entrepreneurship Training Program (No.202611647015), 2026 Provincial-level College Students' innovation and Entrepreneurship Training Program (No. S202611647088S), the First-Class Artificial Intelligence Core Technology Course Construction Project of Communication University of Zhejiang.

References

[1] Afzaal M, Usman M, Fong A C M. Predictive aspect-based sentiment classification of online tourist reviews. Journal of Information Science, 2019, 45(3): 341-363. DOI: 10.1177/0165551518789872.

[2] Li X, Zhang Y, Mei L. Analyzing online reviews of foreign tourists to destination attractions in China: A novel text mining approach. Asia Pacific Journal of Tourism Research, 2023, 28(7): 647-666. DOI: 10.1080/10941665.2023.2255315.

[3] State Council of the People's Republic of China. The 14th Five-Year Plan for Tourism Development. Beijing: State Council of the People's Republic of China, 2021.

[4] Yu Fei, Wang Ru. Research on the Impact of Generative Artificial Intelligence on Tourism Information Adoption Behavior: An Integrated Perspective of TTF and TPB Journal of Mudanjiang Normal University (Social Sciences Edition), 2026(2): 64-73. DOI: 10.13815/j. cnki. Jmtc (pss). 2026.02.012.

[5] Tian Yuhan. Research on the Reconstruction of the Cultural Tourism Digital Content Production and Dissemination Ecosystem Driven by AIGC. Exhibition Economy, 2026(15): 66-69. DOI: 10.19995/j.cnki.CN10-1617/F7.2026.15.066.

[6] Xu Chun, Liu Peizhen, Yan Rong. A Named Entity Recognition Method for User-Generated Content in the Tourism Field. Modern Electronic Technology, 2026, 49(10): 22-28. DOI: 10.16652/j.issn.1004-373x.2026.10.004.

[7] Mckinsey & Company. The promise of travel in the age of AI. New York: McKinsey & Company, 2023.

[8] Zhou Yimeng, Zheng Yuheng, Huang Liang. Research on the Value Co-creation Behavior of Tourists in Inbound Tourism Digital Platforms Empowered by Large Language Models. Journal of Tourism Management, 2026(4): 21-31. DOI: 10.26928/j.cnki.lyxb.2026.04.003.

[9] Li Hua, Kong Jiao, Qiao Zhengyuan, Bai Le. Research on the Impact of Risk Perception on Unsafe Behaviors of Dense Crowds in Urban Scenic Spots. Journal of Safety and Environment, 2023, 23(7): 2379-2387. DOI: 10.13637/j.issn.1009-6094.2022.0463.

[10] Wang Xuejiao, Wang Xiaofeng. The Impact of Negative Online Interaction Information on the Purchase Intention of Heterogeneous Tourism Consumers. Resource Development and Market, 2024, 40(2): 161-171.

[11] Zhang Min, Liu Xuling, Peng Yashi, et al. Research on Image Perception of Tourist Destinations Based on Network Text Analysis Method: A Case Study of Mount Tai Scenic Area Journal of Taishan University, 2022, 44(4): 85-96.

[12] Ren Xiang, Yang Xi, Zhang Junkun, Chen Yao. Tourism Review and Sentiment Analysis of Panzhihua. Modern Information Technology, 2023, 7(24): 131-134. DOI: 10.19850/j.cnki.2096-4706.2023.24.027.

[13] Liu Yanan, Deng Lijie, Song Yang, et al. Research on the Impact of Structured Reviews on Consumers' Willingness to Accept Reviews. Journal of Information Systems, 2022(2): 45-61.

[14] World Tourism Organization (UNWTO), Celth, Breda University of Applied Sciences, et al. “Overtourism”? Understanding and managing urban tourism growth beyond

 Copyright @ STEMM Institute Press

perceptions. Madrid: UNWTO, 2018. DOI: 10.18111/9789284419999.

[15]Li Hui, Zhang Mengying. Application Scenarios, Risks and Challenges, and Enhancement Strategies of Generative Artificial Intelligence Empowering the High-Quality Development of the Cultural and Tourism Industry Dongyue Forum Collection, 2025, 46(12): 98-111+192. DOI: 10.15981/j.cnki.dongyueluncong.2025.12.009.

[16]Zhang M, Yi F, Gursoy D. The effects of generative artificial intelligence on consumers in hospitality and tourism: A systematic review and future research directions. International Journal of Hospitality Management, 2026, 133: 104452. DOI: 10.1016/j.ijhm.2025.104452.

Copyright @ STEMM Institute Press