Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Governing LLM decision authority in tourism: A bounded generative agency framework for multi-stakeholder itinerary planning

Xinping Pu
Sichuan Agricultural University
Xiaoli Xia
University of Hong Kong
Peiyin Li
Sichuan Agricultural University
Yucheng Zhu
University of Hong Kong
Lu Gan
Sichuan Agricultural University
Jialiu Chen
Chongqing University
Yuzhi Zhou
Sichuan Agricultural University
Liangqiang Li
Sichuan Agricultural University
Yunzhong Cao
caoyz@sicau.edu.cn

Sichuan Agricultural University

Research Article

Keywords:

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11706125/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Governing LLM decision authority in tourism: A bounded generative agency framework for multi-stakeholder itinerary planning

Abstract

Large language models (LLMs) increasingly support tourism itinerary decisions, yet factual grounding alone cannot establish whether a recommendation is acceptable to stakeholders or feasible to implement. This study develops SilverRoute, a framework integrating LLMs, geographic information systems (GIS), and multi-criteria decision analysis. Under bounded generative agency, the LLM interprets needs and explains outcomes, while spatial evidence and formal rules govern feasibility, stakeholder acceptability, coordination, and recommendation or abstention. Evaluation used 4,750 facility and service features, 5,459 road features, and 698 cultural points of interest across 48 age-friendly tourism planning scenarios in Ya'an, China. An evidence-grounded LLM achieved 100% factual-claim accuracy but only 45.14% protocol alignment and recommended routes in all eight infeasible cases. Bounding its authority raised alignment to 100% and eliminated inappropriate and forced recommendations. In 40 recommendable scenarios, SilverRoute respected all stakeholder acceptability boundaries, whereas scalar baselines crossed at least one boundary in 17.5–20.0% of cases. Ablations showed that screening acceptability before aggregation provided most of the protection, while Nash bargaining became more consequential as preferences diverged. Time-boundary and repair tests showed that abstention can identify the requirement blocking action and support planner-confirmed adjustments without relaxing protected thresholds. The findings show how explicitly allocating decision rights can make itinerary planning auditable while supporting personalization and protecting stakeholder requirements.

Keywords: Large language models, Generative AI, AI governance, Multi-stakeholder decision-making, Itinerary planning, Age-friendly tourism

1 Introduction

Large language models (LLMs) are moving tourism itinerary planning from information assistance toward consequential decision support. They can interpret open-ended requests, integrate heterogeneous evidence, and explain recommendations in natural language (Hsu et al. 2024; Wong et al. 2023; Fouad et al. 2024; Tussyadiah 2020). This shift is especially important in organized senior tourism, where cultural interests,

health and mobility needs, time budgets, and service constraints are translated into deliverable tour products (Pan et al. 2025; Ali et al. 2026; Tao et al. 2024). Tourism planning therefore involves stakeholders with different interests and responsibilities (Sautter and Leisen 1999; Font and Genzbigelytė-Venturi 2026; Generosi et al. 2025). Its legitimacy as decision support depends on personalization and three governance conditions: stakeholder representation, the allocation of decision authority, and action eligibility. Together, these conditions make responsible LLM-enabled itinerary planning an important tourism research problem.

The first condition becomes concrete when professionals plan routes for older groups. The intended users of itinerary-planning systems for older travelers include travel-agency senior-tour product managers, local tourism and destination-management authorities, and organizations serving older adults. They plan or govern routes for groups with different health and mobility profiles. Explicit representation is therefore important for older travelers' needs for medical support, manageable walking and activity, rest, comfort, and cultural experience during early product design (Li et al. 2024; Darcy and Dickson 2009; Tang et al. 2025). At the same time, destination stewards consider public safety, cultural-resource use, and regional coordination, whereas tourism operators emphasize product development and practical delivery (Hall 2011; Yao et al. 2025; Trilaksono et al. 2025). Starting from one actor's priorities can therefore produce an attractive route for one party while reducing its acceptability to another (Bramwell and Lane 2011). The resulting challenge is to preserve distinct stakeholder claims and coordinate conflicting interests so that each actor's minimum requirements retain decision standing.

Trustworthy LLM-mediated planning requires both stakeholder representation and reliable evidence. Current efforts to improve tourism LLMs mainly address hallucination, misinformation, and factual grounding (Dwivedi et al. 2024; Kim et al. 2025b; Christensen et al. 2024; Kim et al. 2024). Geographic and tourism-domain knowledge representations can strengthen spatial grounding in pretrained models (Ramrakhiyani et al. 2025; Fan and Chen 2024). In operational planning, retrieval, GIS, and application programming interfaces can then connect model inputs to current attraction, road, travel-time, weather, and service evidence. Additionally, safety-critical LLM research combines evidence grounding with stepwise verification and uncertainty control (Lu et al. 2026). Factual accuracy therefore needs a complementary governance rule because model outputs may also influence constraints, stakeholder priorities, and recommendation status. Accordingly, the governance question is how such outputs should affect a consequential tourism decision.

Integrating GIS with multicriteria decision analysis provides reproducible and spatially explicit analysis (Malczewski 2006; Jankowski 1995), while generative AI can make requirements and spatial evidence easier to interpret. Their integration therefore requires explicit authority allocation. Human oversight and post-hoc explanation can improve supervision, while stage-specific decision rights specify which itinerary elements an LLM can propose, which require confirmation, and which are determined by validated spatial and decision procedures (Shrestha et al. 2019; Raisch and Krakowski 2021; Baird and Maruping 2021; Fügener et al. 2022; Berente et al. 2021). Additionally, recent agentic-AI evidence shows that autonomy level and user control shape trust and

workload (Geninatti Cossatin et al. 2026). Consequently, the division of decision rights between generative AI and auditable GIS-based multicriteria procedures remains an important question.

The allocation of decision rights leads to a further question about action eligibility. Real planning requests can produce an empty feasible set under time limits, mobility needs, stakeholder thresholds, and plausible operating conditions. Itinerary recommendation systems commonly optimize or rank available alternatives by relative performance (Borràs et al. 2014; Otaki and Baba 2025; Chang et al. 2025; Ding et al. 2024; Xu et al. 2024; Dalla Vecchia et al. 2024; Dietz et al. 2025). Related high-stakes ranking research shows that fair exposure and fair downstream outcomes can diverge (Fabris et al. 2026). Similarly, the highest-ranked itinerary may remain ineligible when it violates confirmed constraints or stakeholder thresholds. Distinguishing the best available itinerary from one that is sufficiently acceptable for recommendation is thus an important problem for responsible tourism AI. Accordingly, it is necessary to determine when withholding a recommendation constitutes an appropriate decision outcome.

Based on these issues, this study addresses three research questions:

(I) Stakeholder coordination. How can LLM-enabled itinerary planning represent the distinct interests and protect the claims of older travelers, destination stewards, and tourism operators while reconciling conflicts as their preferences diverge?

(II) Generative decision authority. How should decision authority be allocated between generative AI and auditable GIS-based multicriteria procedures so that requirement interpretation and evidence communication remain flexible while formal route selection remains deterministic, spatially traceable, and auditable? Does an ex ante authority boundary improve protocol-aligned action beyond factual grounding alone?

(III) Responsible recommendation. Highly ranked candidates may still violate time or mobility constraints or fall below stakeholder thresholds, including those for schedule robustness. When should the system withhold a recommendation and responsibly abstain?

To answer these questions, the study develops SilverRoute, an LLM–GIS framework with multicriteria decision analysis governed by bounded generative agency for tourism itinerary planning for older travelers. Under bounded generative agency, the LLM interprets natural-language requirements and explains computed results, whereas a confirmed decision contract and an auditable GIS-based multicriteria core govern formal decisions. The core uses road-constrained beam search to generate candidate itineraries and combines reliability-adjusted Criteria Importance Through Intercriteria Correlation weights with the three stakeholders' subjective weights to calculate separate utilities. Hard constraints, stakeholder vetoes, and reservation utilities then establish the agreement set before weighted Nash bargaining selects a compromise. Additionally, schedule-stress scenarios assess operational robustness; an empty agreement set triggers responsible abstention. Finally, the Agency Boundary Experiment, conflict tests, multicriteria benchmarks, and sensitivity analyses evaluate authority allocation, stakeholder coordination, action eligibility, and robustness. Editable QGIS outputs preserve spatial traceability.

2 Literature review

2.1 Stakeholder-aware personalization

Tourism recommendation has moved from preference matching toward contextual and conversational personalization. Earlier work addressed information overload and heterogeneous, changing preferences (Gretzel and Fesenmaier 2006; Jeng and Fesenmaier 2002; Park and Fesenmaier 2014; Borràs et al. 2014). More recently, generative AI has extended this trajectory by interpreting open-ended requests and producing flexible itineraries; personalization may also improve recommendation acceptance (Kim et al. 2025a; Wang et al. 2025; Abou-Shouk et al. 2024; Friedli et al. 2025).

Implementation makes personalization multi-stakeholder. A route is simultaneously a traveler experience, an operator-delivered product, and a spatial intervention. Senior tourism makes this interdependence salient because older travelers differ in mobility, health, fatigue, rest, and service needs (Nicolau et al. 2020; Pan et al. 2025; Ali et al. 2026; Leung et al. 2025). Moreover, planners may act on travelers' behalf, so the beneficiary and system operator can be different actors. Tourism planning therefore requires explicit stakeholder representation (Sautter and Leisen 1999; Font and Genzbigelytė-Venturi 2026).

Current personalization still tends to treat the initiating user as the principal optimization target (Abdollahpouri et al. 2020). Meanwhile, recommender-system research increasingly represents fairness across stakeholders and allows users to control multiple objectives (Amigó et al. 2023; Dokoupil et al. 2025; Banerjee et al. 2025; Delic et al. 2018). Under conflict, one party's gains can still obscure another's unacceptable losses. Distinct stakeholder claims therefore need explicit protection, while authority allocation must specify which tasks generative AI may perform and which decisions remain auditable.

2.2 Generative decision authority

Tourism research first emphasized generative AI for conversational planning and knowledge support (Wong et al. 2023; Hsu et al. 2024; Tussyadiah 2020; Fouad et al. 2024; Cruz et al. 2025), then increasingly addressed trust, error, bias, and responsible deployment (Kim et al. 2025b; Huang et al. 2026; Christensen et al. 2024; Jia et al. 2026; Foroughi et al. 2025; Zhao et al. 2026; Wong et al. 2025; Glikson and Woolley 2020). Geographic and tourism-domain knowledge representations can improve the evidence available to pretrained models (Ramrakhiyani et al. 2025; Fan and Chen 2024). Additionally, authority allocation determines who may revise constraints, priorities, or recommendations. Epistemic reliability and decision authority are therefore distinct.

Agency literature supports this distinction. Bounded agency situates action within structures and resources (Evans 2007; Aaltonen 2013), while generative-AI research examines how design allocates agency between humans and models (Satyanarayan and Jones 2024). Additionally, AI can redistribute organizational decision rights (Raisch and Krakowski 2021; Shrestha et al. 2019; Baird and Maruping 2021; Fügener et al. 2022; Berente et al. 2021), and recent agentic-AI evidence links autonomy settings to user control, trust, and workload (Geninatti Cossatin et al. 2026). Here, generative

agency denotes model discretion to construct representations that affect the choice set, evaluation, or formal action. Bounded generative agency constrains that discretion ex ante to authorized stages.

Accordingly, interpretation and explanation can remain flexible while spatial measurement, stakeholder weights, vetoes, bargaining, and recommend-or-abstain actions remain protocol-governed. Stage-specific decision rights specify who may change what before action. This sharpens debates about AI support versus substitution (Mikulić 2025). Moreover, explanation can affect decision performance and behavioral trust, while formal decision rights require separate specification (Wang and Ding 2024). The key empirical question is whether an ex ante authority boundary improves protocol-aligned action beyond factual grounding.

2.3 Recommendation legitimacy

Multicriteria decision support compares alternatives across heterogeneous criteria. Simple additive weighting produces a scalar score; the technique for order preference by similarity to an ideal solution ranks distance from ideal solutions (Hwang and Yoon 1981), while multicriteria optimization and compromise solution seeks compromise through group utility and regret (Opricovic and Tzeng 2004). Compensatory aggregation can let strong performance offset critical weakness (Bouyssou 1986; Roy 1991; Munda 2004).

For tourism, a single score can therefore conceal whose loss is offset by whose gain (Keeney and Raiffa 1976). Multi-objective recommendation often balances competing objectives through explicit trade-offs (Khaitan and Shrivastava 2026). Accordingly, a route may remain unacceptable when it violates an older traveler's health or fatigue threshold despite high cultural value or operator efficiency. Building on non-compensatory traditions (Roy 1991; Munda 2004), acceptability before aggregation checks hard feasibility, vetoes, and stakeholder reservation utilities before forming an agreement set. Weighted Nash bargaining then coordinates gains among jointly acceptable routes (Nash 1950). Legitimacy is thus procedural rather than purely rank-based.

Relative superiority also differs from actionability. Related high-stakes ranking research shows that fair exposure and fair downstream outcomes can diverge (Fabris et al. 2026). Similarly, the best available route may remain fragile or unacceptable. Reject-option research formalizes withholding outputs when reliability is insufficient (Chow 1970; Hendrickx et al. 2024; Hallberg Szabadváry et al. 2025). Responsible abstention extends this logic to protocol eligibility: an empty protected agreement set makes withholding recommendation the legitimate outcome.

3 Methodology

3.1 Study area and data

3.1.1 Study area

The study focuses on Yucheng District, Mingshan District, Lushan County, and Tianquan County in Ya'an City, Sichuan Province, China (Fig. 1). The area links the

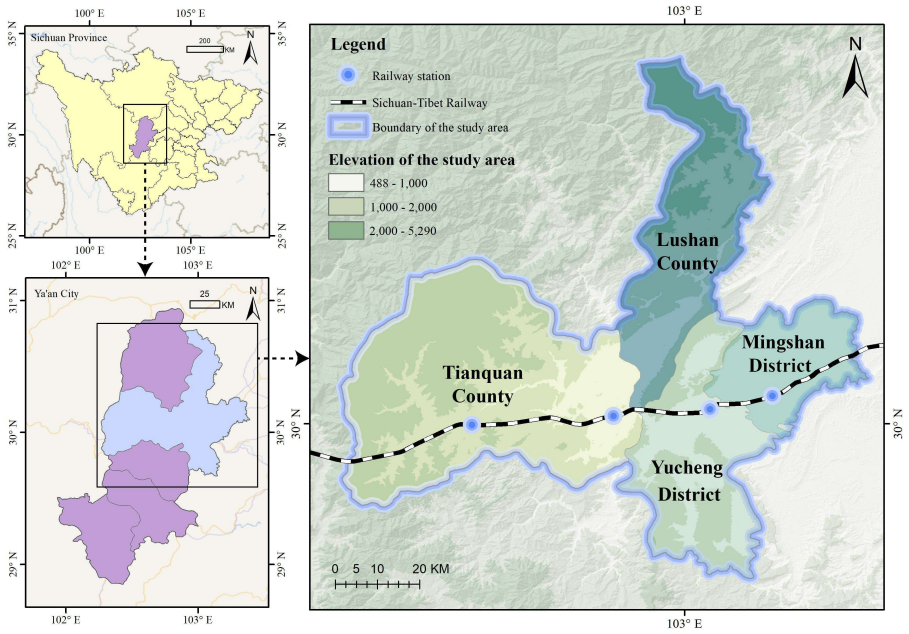


Fig. 1 Study area and spatial context (QGIS 3.40.8). The four Ya'an jurisdictions provide an urban-peripheral gradient for within-region route-planning experiments

Chengdu metropolitan market with resource-rich mountain destinations and presents a clear urban-peripheral gradient. Yucheng and Mingshan have stronger urban and service functions, whereas Lushan and Tianquan represent more peripheral settings. Consequently, the four jurisdictions provide meaningful variation in travel time, service support, cultural opportunities, and route feasibility within one regional context. This structure supports within-region tests across varied local settings.

3.1.2 Data and evaluation framework

The spatial evidence base integrated 21 criterion rasters and 14 vector layers. Specifically, 10 facility and service layers contained 4,750 features, while the transport network comprised 5,459 road features, 156 railway-line features, and four stations. Additionally, 698 cultural points of interest supported route-candidate screening. Together, these data represented cultural, health, mobility, age-friendly, service, and destination conditions. Additionally, all indicators were transformed to a higher-is-better orientation and robustly scaled to [0,1] using the 5th and 95th percentiles. Detailed preprocessing and provenance are reported in Supplementary Section S1.

The evaluation framework follows a two-level multiattribute structure. Level-1 dimensions define the eight route criteria used in stakeholder utility, while Level-2 indicators operationalize each dimension through GIS evidence or route calculations (Table 1). This hierarchy follows multiattribute utility theory by keeping decision attributes explicit before aggregation (Keeney and Raiffa 1976). Its substantive dimensions also reflect age-friendly and multi-stakeholder requirements (Darcy and Dickson 2009; Pan

et al. 2025; Ali et al. 2026; Sautter and Leisen 1999). Aggregation rules are reported in Supplementary Table S1 (Online Resource 1). Additionally, schedule robustness and route efficiency are calculated after route construction.

Level-2 indicators provide the evidence inputs, and some spatial inputs support more than one Level-1 dimension. Schedule robustness and route efficiency are derived at the route level. The eight Level-1 dimensions enter stakeholder utility.

3.2 Decision framework

3.2.1 Bounded generative agency and decision contract

Building on the bounded-agency and generative-agency foundations in Section 2.2, bounded generative agency is operationalized as an ex ante allocation of decision rights. This structure fits itinerary planning because natural-language interpretation benefits from generative flexibility, whereas route feasibility, stakeholder preferences, and action eligibility require auditable control. Let G denote the generative model, q the natural-language request, X the verified evidence package, θ the locked decision parameters, and F the deterministic GIS-based multicriteria planner. The LLM proposes a structured contract $\hat{c} = G(q, X)$, and formal action is computed from the confirmed canonical contract c:

$$a = F(c, X, \theta), \qquad e = G(q, X, a), \tag{1}$$

where a is the recommend-or-abstain action and e is an optional explanation. The equations provide an operational representation of the authority boundary. The canonical contract, weights, constraints, utilities, bargaining, and final action remain governed by F. Accordingly, the authority architecture is tested by comparing identical scenario families under different allocations of LLM authority.

Fig. 2 and Supplementary Table S6 summarize this stage-specific allocation. The interface converts natural-language requests into planner-confirmed decision fields before formal evaluation: the LLM proposes a structured contract, whereas the formal planner uses the confirmed contract. This keeps linguistic interpretation separate from formal decision variables. For example, "less walking" can be interpreted as a salient mobility preference, while the formal walking threshold comes from the confirmed contract. Missing fields are resolved through explicit confirmation or recorded experimental defaults, distinguishing requested, assumed, and inferred information. The deployed interface makes this sequence visible to the planner.

3.2.2 Route generation

After semantic filtering, candidate cultural points of interest inherited the six evidence dimensions defined in Section 3.1.2. Candidate preselection preserves multi-stakeholder relevance through a recall score that combines cross-stakeholder mean relevance with the strongest stakeholder-specific relevance:

$$S_i^{recall} = 0.70\overline{S}_i + 0.30 \max_s(S_{i,s}) + B_i, \tag{2}$$

where B_i represents auditable bonuses for request relevance and short-trip proximity. Stakeholder-specific recall tranches additionally retain alternatives with strong

Table 1 Two-level evaluation system for route assessment

Level-1 dimension	Level-2 indicator or operationalization	Decision role
Cultural experience	Cultural value	Experience quality
	Cultural facility service provision	
	Cultural resource accessibility	
	Cultural agglomeration	
	Cultural appeal to older adults	
Health and safety	Medical service accessibility	Protected well-being
	Cross-regional medical support	
	Slope suitability	
	Distance from rivers	
	Distance from historical seismic activity	
	Distance from recorded geohazards	
Mobility access	Railway-station accessibility	Travel accessibility
	Road-network density	
	Source-market accessibility	
Age-friendly comfort	Age-friendliness of cultural attractions	Physical burden and comfort
	Cultural appeal to older adults	
	Eldercare-facility agglomeration	
	Medical service accessibility	
Service convenience	Shopping services	Daily and operational support
	Financial services	
	Public services	
Schedule robustness	Completion under predefined delay/stress scenarios	Operational reliability
Destination synergy	Tertiary-industry context	Destination coordination
	Cultural-corridor agglomeration	
	Eldercare-facility agglomeration	
	Source-market accessibility	
Route efficiency	Route distance	Delivery efficiency
	Remaining time budget	

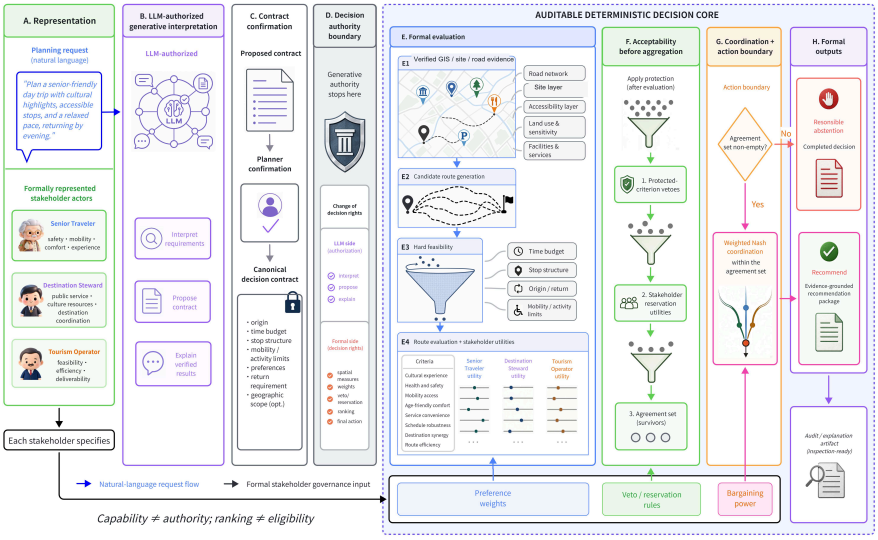


Fig. 2 SilverRoute decision architecture. Generative interpretation is separated from formal evaluation and action. Formal authorization governs decision authority and action eligibility, while generative capability supports interpretation and explanation

stakeholder-specific relevance. The baseline search used a 220-site preselection cap, a 48-site candidate pool, a beam width of 300, and at most 120 final route candidates; complete settings are reported in Supplementary Table S4.

Routes were generated on a noded road graph using shortest-path distances. Where road-network topology left a site pair disconnected, the planner used a penalized Euclidean fallback and recorded the event as a provenance flag. Beam search controlled the combinatorial route space. Additionally, wider-search and reduced-instance exact-enumeration checks assessed search adequacy (Supplementary Section S5).

Route duration combined travel time, visit time, age-friendly transition allowance, and rest/meal time. Schedule robustness was evaluated under predefined typical, mild-delay, peak/rain, adverse, and severe-stress scenarios (Supplementary Table S5). All distances, travel times, robustness scores, and route-efficiency values were computed by deterministic routines. Route efficiency combined distance and time reserve:

$$E(r) = 0.65 \exp(-D_r/70) + 0.35 \max\left(0, 1 - T_r/T_{budget}\right). \tag{3}$$

3.2.3 Stakeholder utility and acceptability

The formal decision represents three stakeholders. The Senior Traveler prioritizes safety, physical burden, age-friendly comfort, and experience. The Destination Steward prioritizes public safety, cultural-resource use, service coordination, and destination interests. The Tourism Operator prioritizes route efficiency, schedule feasibility, service

support, and deliverability. These roles reflect established stakeholder and destination-governance perspectives (Sautter and Leisen 1999; Hall 2011; Font and Genzbigelytė-Venturi 2026). Each stakeholder enters the model as an auditable preference-and-constraint profile. Their interaction occurs through separate route evaluation, a shared agreement set, and bargaining within that set. Full weight vectors and protection parameters are reported in Supplementary Tables S2–S3.

The stakeholder-specific subjective weights $w_{s,j}^{S}$ were constructed as research-design priors from documented role priorities and normalized to one. Baseline bargaining powers were 0.50 for the Senior Traveler, 0.28 for the Destination Steward, and 0.22 for the Tourism Operator. Sensitivity analysis evaluated the influence of both preference weights and bargaining power.

For the hard-feasible route set, the eight benefit-oriented route criteria formed a normalized matrix $Z = [z_{ij}]$. Criteria Importance Through Intercriteria Correlation weighting estimates the objective discriminating information carried by each criterion (Diakoulaki et al. 1995). For criterion j, it combines dispersion and inter-criterion conflict:

$$C_j = \sigma_j \sum_k (1 - r_{jk}), \qquad w_j^C = \frac{C_j}{\sum_l C_l}, \tag{4}$$

where σ_j is the criterion's standard deviation and r_{jk} is its Pearson correlation with criterion k. Evidence-reliability coefficients then adjust and renormalize the raw objective weights:

$$\widetilde{w}_j^C = \frac{\rho_j w_j^C}{\sum_l \rho_l w_l^C}. \tag{5}$$

Objective information received a baseline share $\lambda = 0.30$, while stakeholder value weights retained 0.70:

$$w_{s,j}^F = (1 - \lambda) w_{s,j}^S + \lambda \widetilde{w}_j^C. \tag{6}$$

Stakeholder utility follows the additive multiattribute form of Keeney and Raiffa (1976):

$$U_s(r) = \sum_j w_{s,j}^F z_j(r). \tag{7}$$

Accordingly, the objective-weighting method contributes objective information, while stakeholder preference weights and bargaining power retain distinct normative roles. Three screens then established the agreement set. First, a route had to satisfy structural hard constraints such as total time, stop count, origin/return requirements, and relevant activity limits. Second, stakeholder-specific protected criteria were subject to a hybrid veto:

$$T_{s,j} = \max\left(a_{s,j}, q_{s,j} z_j^{ideal}\right), \tag{8}$$

where $a_{s,j}$ is an absolute floor and $q_{s,j}$ is a relative threshold applied to the criterion ideal. Third, each stakeholder required utility above a reservation point:

$$d_s = \eta_s U_s^{ideal}, \qquad U_s^{ideal} = \max_r U_s(r). \tag{9}$$

The reservation ratios were 0.75 for the Senior Traveler and 0.60 for the Destination Steward and Tourism Operator. Together, the three screens preserve protected minima before aggregation. Membership in the resulting set is termed agreement eligibility; protocol-aligned action evaluates whether the final recommend-or-abstain decision matches that formal status.

3.2.4 Nash coordination and responsible abstention

Routes satisfying all hard constraints, veto thresholds, and stakeholder reservation utilities entered bargaining. For each eligible route, stakeholder gain above the reservation point was normalized by the feasible ideal range:

$$\widetilde{g}_s(r) = \frac{U_s(r) - d_s}{U_s^{ideal} - d_s}. \tag{10}$$

The selected route maximized the weighted log Nash product:

$$\max_r \sum_s \alpha_s \log\left[\max(\widetilde{g}_s(r), \epsilon)\right], \tag{11}$$

where α_s denotes normalized bargaining power. The bargaining layer therefore coordinates gains among already acceptable alternatives, with criterion weights and stakeholder boundaries fixed upstream. This log formulation is monotonically equivalent to the classical weighted Nash product (Nash 1950), while normalization by attainable surplus places stakeholder gains on a common scale. An empty agreement set returns `completed_infeasible`, operationalized as responsible abstention, and records a completed decision process at the action boundary.

3.3 Validation strategy

The validation program follows the three research questions. Stakeholder coordination was examined through formal multi-stakeholder decisions, traditional multicriteria comparisons, mechanism ablations, and conflict stress. Generative authority was tested in the Agency Boundary Experiment. Recommendation legitimacy was examined through recommend-versus-abstain behavior, threshold and time-boundary tests, and robustness analysis. Additionally, within-area transportability, age-friendly proxy tests, search validity, and fallback sensitivity assessed scope and implementation. Detailed designs are reported in Supplementary Section S5.

3.3.1 Core validation

Formal factorial validation. Forty-eight scenarios varied origin, trip budget, stop structure, and traveler requirements. The purpose was to test whether the system could complete the formal decision protocol across heterogeneous itinerary conditions and correctly separate recommendation from abstention.

Agency Boundary Experiment. This primary theory test used the same 48-scenario family under four conditions: (1) deterministic reference; (2) unbounded LLM discretion;

(3) evidence-grounded LLM discretion over formal action; and (4) bounded generative agency. In the fourth condition, the canonical contract governed candidate scope, constraints, utility, veto, reservation, bargaining, and action. All LLM conditions used DeepSeek-V4-Pro (`deepseek-v4-pro`) with temperature = 0.2 and identical generation settings. The deterministic condition contained 48 observations. Each LLM condition contained three persisted realizations per scenario, yielding 480 observations in total.

Primary outcomes were protocol-aligned action, hard-constraint satisfaction, agreement eligibility, inappropriate recommendation, forced recommendation conditional on protocol infeasibility, semantic-contract accuracy, unsupported factual claims, and route stability. Semantic-contract accuracy is the share of 11 interpretation fields matching the scenario contract. Factual-claim accuracy is the share of context-supported claims matching verified evidence, with a 2% tolerance for numeric values. Route stability is the within-scenario share of repeated realizations matching the modal decision-and-route signature. For paired bounded-authority versus evidence-grounded inference, repeated LLM observations were first averaged within scenario-condition, making the scenario the independent unit. Effects used 10,000 bootstrap resamples for 95% confidence intervals and 20,000 paired randomization iterations for p values (analysis seed 20260808).

Traditional multicriteria comparison. Weighted overlay, the ideal-solution method, and the compromise-solution method received the same hard-feasible route matrices and criterion information as SilverRoute. Outcomes included agreement eligibility, selection agreement, minimum stakeholder utility, mean utility, and utility dispersion. Paired differences used two-sided Wilcoxon signed-rank tests.

Mechanism and search validity. Additional ablations held candidate routes and weights fixed while varying veto, reservation, and acceptability gates and substituting mean utility or maximin. Reduced instances were also checked against exact enumeration and classic heuristic search methods. These analyses isolate mechanism effects and assess beam-search adequacy.

3.3.2 Robustness and scope tests

Conflict stress. To test when Nash coordination becomes decision-relevant, each stakeholder's final preference vector was sharpened with the same exponent while preserving preference direction:

$$w_{s,j}^{(\gamma)} = \frac{(w_{s,j})^{\gamma}}{\sum_k (w_{s,k})^{\gamma}}, \qquad \gamma \in \{1, 1.5, 2, 3\}. \tag{12}$$

Selection divergence from mean utility, Senior-Traveler-only choice, and maximin was measured as conflict increased.

Sensitivity and Monte Carlo analysis. Search depth, visit/transition time assumptions, bargaining power, reservation ratios, veto levels, objective-weighting method and share, disconnected-network fallback penalties, and bounded preference uncertainty were varied. A 500-draw Monte Carlo analysis perturbed governance parameters over 40 formally recommendable scenarios, producing 20,000 protocol evaluations. Separate threshold- and time-boundary analyses located recommendation-to-abstention transitions and tested counterfactual repairs while retaining protected stakeholder

thresholds. The time-boundary sweep recorded hard infeasibility separately from completed abstention; both result in abstention at the action boundary.

Within-area transportability. Sixteen county-constrained scenarios tested whether the same protocol remained operational within the four Ya'an jurisdictions.

Age-friendly proxy validation. Twelve wheelchair/mobility-aid proxy scenarios assessed differentiated screening from the available GIS evidence.

3.4 Spatial audit and implementation

For spatial audit, QGIS 3.40.8 linked candidate and selected routes to study-area, transport, facility, stop, and origin layers. Additionally, the user interface supports contract confirmation and result inspection, while the confirmed contract and deterministic planner govern ranking and action. Supplementary Table S4 reports implementation parameters, and Supplementary Table S7 provides the validation registry.

4 Results

The results are organized into three claim domains. Section 4.1 establishes formal action outcomes and tests bounded generative authority. Section 4.2 examines stakeholder protection and coordination. Section 4.3 evaluates robustness and validation.

4.1 Protocol outcomes and generative authority

4.1.1 Formal decision outcomes

All 48 formal scenarios reached a valid protocol outcome. Forty produced a protocol-eligible recommendation, while eight ended in responsible abstention because the agreement set was empty. Thus, the experiment shows that SilverRoute distinguishes between cases in which action is supportable and cases in which withholding a recommendation is the protocol-consistent outcome.

Boundary analysis further clarified this behavior. Across 52 time-boundary cases, recommendation status changed as the available schedule varied. Additionally, all eight minimum-repair cases recovered a protocol-eligible option through explicit time-budget or stop-count adjustment while preserving protected stakeholder thresholds.

Importantly, every issued recommendation satisfied the formal hard constraints, retained complete decision-trace evidence coverage, and was Pareto-efficient within the evaluated feasible set. Across the objective-weight sensitivity grid, exact-route stability reached 91.25%. Mean pairwise Spearman rank correlation among stakeholder rankings was 0.9309, showing substantial common ground in the baseline scenarios. Supplementary Table S8 summarizes these outcomes. This high baseline agreement provides the reference point for the later bargaining tests: when stakeholder rankings already converge, similar choices across coordination rules are expected, so the conflict experiment tests when Nash coordination becomes decision-relevant.

Table 2 Traditional multicriteria comparison

Method	Agreement-eligible	Selection agreement with SilverRoute	Mean minimum utility	Mean stakeholder utility	Mean utility dispersion
SilverRoute (bounded authority + Nash)	**1.000**	1.000	0.54697	0.56000	0.01155
Weighted overlay	0.800	0.800	**0.54884**	**0.56318**	0.01231
Ideal-solution method	0.825	0.350	0.54337	0.55522	**0.01020**
Compromise-solution method	0.800	0.400	0.54354	0.55712	0.01205

4.1.2 Agency Boundary Experiment

The Agency Boundary Experiment provides the central test of bounded generative agency (Fig. 3). The unbounded LLM achieved 25.7% protocol alignment (Fig. 3a). Additionally, evidence grounding achieved 100% factual-claim accuracy but still forced a recommendation in all eight infeasible cases, whereas bounded authority restored protocol-aligned action and eliminated inappropriate and forced recommendations. All four paired bounded-authority effects excluded zero: protocol alignment, avoidance of inappropriate recommendation, and semantic-contract accuracy each had p ¡ .001, while avoidance of forced recommendation had $p = .008$ (Fig. 3b). Thus, factual grounding improved epistemic reliability, whereas the authority boundary governed action eligibility.

4.2 Stakeholder protection and coordination

4.2.1 Multicriteria comparison and mechanism ablation

The traditional multicriteria comparison separates protocol eligibility from scalar utility performance. All 40 SilverRoute recommendations were agreement-eligible. By comparison, agreement eligibility was 80.0% for weighted overlay, 82.5% for the ideal-solution method, and 80.0% for the compromise-solution method. Thus, 17.5–20.0% of the scalar baseline optima crossed at least one stakeholder's complete acceptability boundary.

Scalar utility differences were small and metric-dependent. Minimum utility was comparable between SilverRoute and weighted overlay ($p = .207$). SilverRoute exceeded the ideal-solution method (p ¡ .001) and the compromise-solution method ($p = .022$). Conversely, weighted overlay achieved slightly higher mean utility, whereas the ideal-solution method produced lower utility dispersion. Table 2 reports the complete comparison.

Together, these results locate SilverRoute's empirical advantage in agreement eligibility, with marginal scalar-score gains secondary. The protocol preserves stakeholder

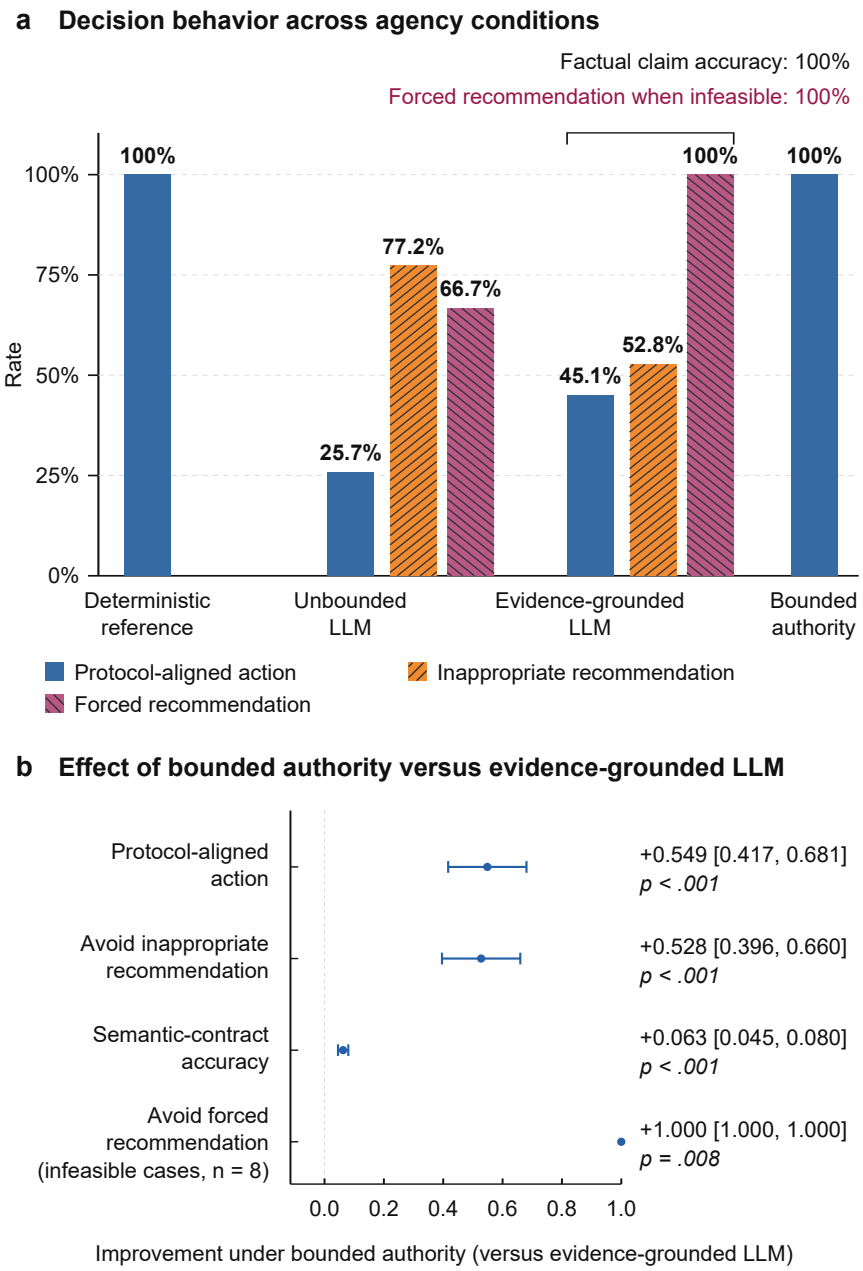


Fig. 3 Agency boundary experiment. Panel a compares decision behavior. Panel b reports bounded-authority effects relative to the evidence-grounded LLM with bootstrap 95% confidence intervals

minima before ranking the remaining alternatives, while the scalar baselines can select a numerically attractive route outside the agreement set.

Mechanism ablation isolated this effect (Fig. 4a). With the full acceptability gate retained, all recommendations remained protocol-eligible. By contrast, removing the veto gate reduced full-protocol eligibility to 68.75%, and removing the full acceptability gate reduced it to 66.67%. Both variants converted the eight reference-abstention cases into recommendations. Thus, acceptability-before-aggregation provided the main protection effect, while Nash coordinated choice within the eligible set.

4.2.2 Stakeholder conflict experiment

In the main formal scenarios, high stakeholder rank correlation led Nash to agree with simpler rules in most cases. Across the 40 recommended scenarios, Nash differed from mean-utility choice in 20.0%, Senior-Traveler-only choice in 32.5%, and maximin in 22.5%. This convergence is consistent with the substantial common ground observed in Section 4.1.1.

Conflict stress then showed when bargaining becomes substantively relevant. At $\gamma = 1$, Nash and mean utility selected the same route in all stress-test scenarios. At $\gamma = 3$, Nash differed from Senior-Traveler-only and maximin choice in 32.5% of scenarios and from mean utility in 15.0% (Fig. 4b). Thus, stronger preference conflict increased the frequency with which the bargaining rule changed the selected compromise. Detailed scalar-utility comparisons are reported in Supplementary Section S5.1.

The pattern is therefore clear: Nash has limited decision impact when stakeholder preferences already converge and becomes more influential as those preferences diverge. Within SilverRoute, its role is to structure agreement over normalized gains after the non-compensatory boundaries have been satisfied.

4.3 Robustness and validation

4.3.1 Sensitivity and responsible abstention

Search-depth tests showed that the final decision was stable even as the candidate set expanded. Across the compact benchmarks, wider search substantially increased candidate recovery while preserving the selected route. Reduced exact enumeration provided a stricter check: beam search matched agreement status in all 12 instances and achieved numerical-scale mean regret (Supplementary Fig. S1).

Governance-parameter sensitivity also behaved consistently with the model design. Moderate changes in bargaining power and reservation ratios preserved agreement and selection in most cases. Tighter veto thresholds had stronger effects because vetoes define the action boundary.

Across 500 Monte Carlo draws and 40 recommendable scenarios, 96.04% of the 20,000 protocol evaluations retained an agreement. Conditional on agreement, selection stability reached 97.91%. These results indicate that moderate joint perturbations rarely changed either action eligibility or the selected route.

Additional stress tests showed the same pattern. Across stakeholder-preference perturbations, recommendation-versus-abstention decisions remained stable, with variation in exact routes (Supplementary Fig. S2). Additionally, disconnected-road penalty

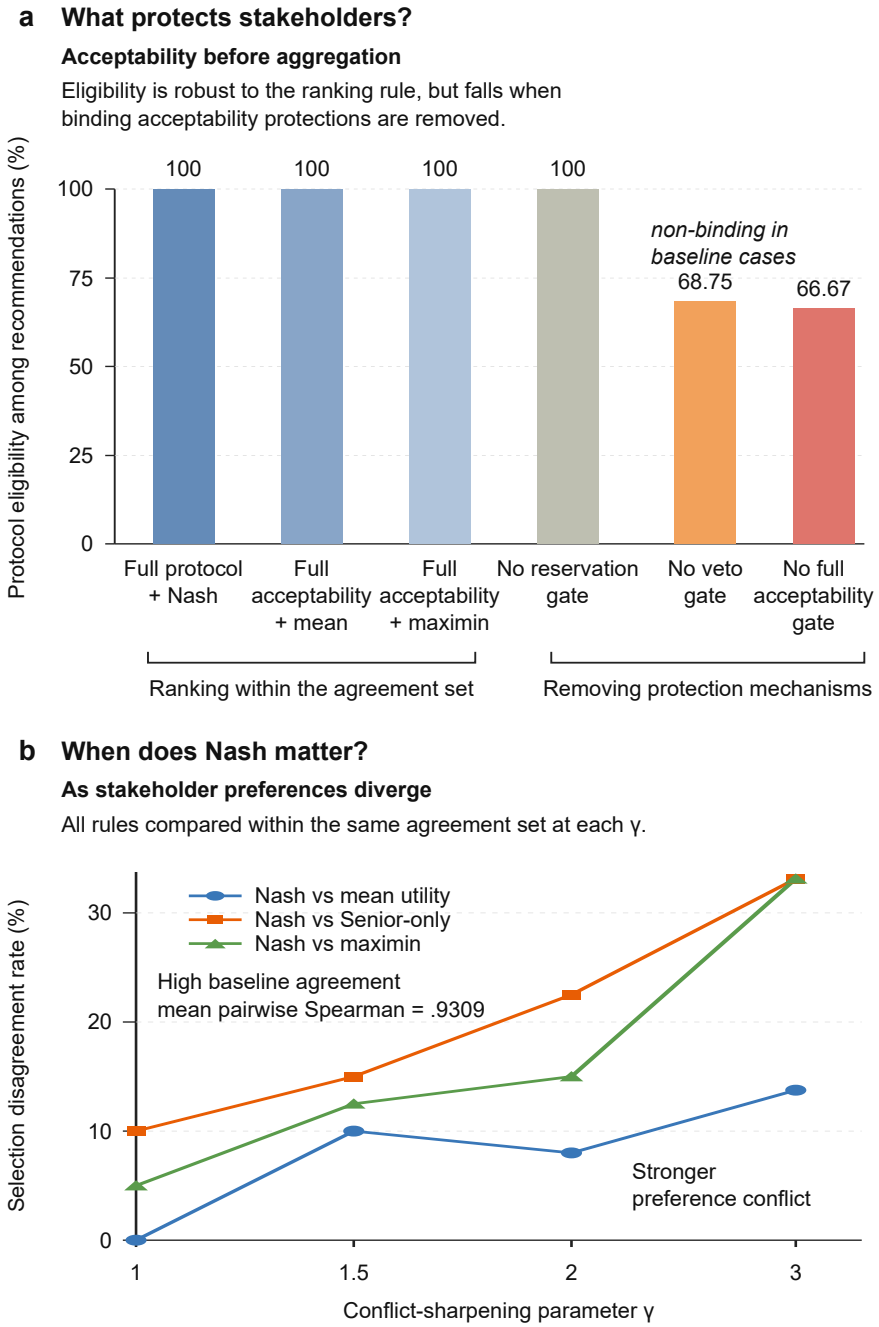


Fig. 4 Stakeholder protection and bargaining relevance. Panel a: protocol eligibility by variant. Panel b: selection disagreement by conflict. Denominators: 40 recommendations for the first four variants and 48 for the last two

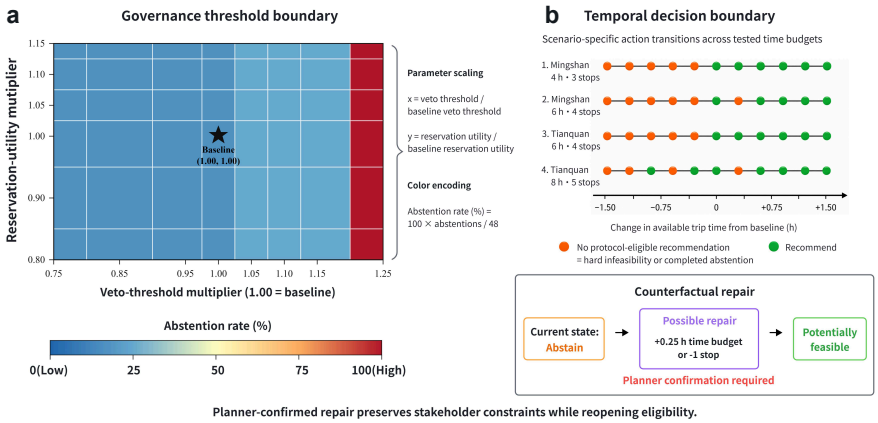


Fig. 5 Responsible abstention boundary. Panel a shows that stricter veto thresholds increased abstention, whereas reservation-utility scaling preserved action rates across the tested grid. Panel b shows the scenario-specific action transitions under changes to the time budget. Planner-confirmed repairs restored eligibility while preserving stakeholder constraints

changes preserved the action in every tested case, while exact-route stability remained high (Supplementary Fig. S3). Thus, action-level conclusions were more stable than exact route selection. Under a deliberately harsher holdout stress of 1.80× travel time and 1.25× visit time, 8 of 40 baseline recommendations remained feasible, defining an extreme operational boundary (Supplementary Section S5.1).

Across the full analysis, nonrecommendation defined both a governance threshold and a temporal decision boundary (Fig. 5). The temporal sweep distinguished hard infeasibility from completed abstention; planner-confirmed changes to time or itinerary structure could restore eligibility.

Across the 48 threshold combinations, all reference-infeasible cases remained abstentions. Recommendation rates remained 83.3% through the baseline veto scale, fell to 75.0% at scales of 1.05–1.15, and reached zero at 1.25. These rates were invariant to reservation scaling.

4.3.2 Transportability and evidence limits

All 16 county-constrained scenarios across Yucheng, Mingshan, Tianquan, and Lushan produced a recommendation with 100% hard-constraint satisfaction, 100% county-scope satisfaction, and 100% Pareto-efficient selection. Exact-route stability was 69.53%, while mean pairwise stakeholder Spearman rank correlation was 0.8601. Thus, protocol-level behavior remained consistent across the four jurisdictions, while route identity adapted to local choice sets.

The age-friendly proxy experiment produced 10 recommendations and two responsible abstentions across 12 wheelchair/mobility-aid scenarios. Among jointly recommendable pairs, route selection remained stable. Thus, the available proxies affected eligibility more clearly than within-set ranking.

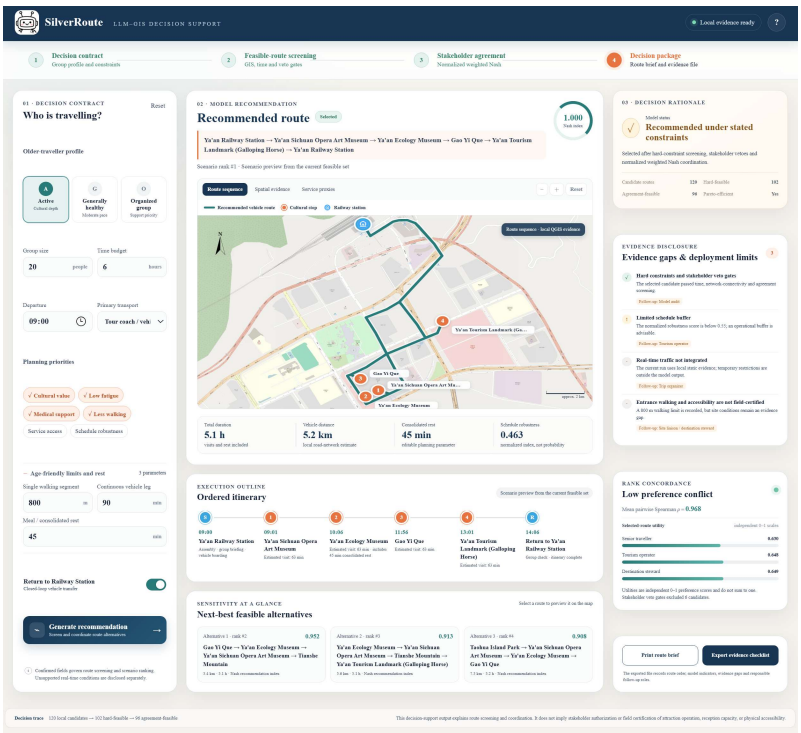


Fig. 6 SilverRoute planning interface. The interface presents planner-confirmed decision fields, the recommended route, stakeholder agreement, and evidence limits

4.3.3 Illustrative case and spatial audit

To demonstrate the audit chain, a six-hour, four-stop, vehicle-supported itinerary was generated for 20 active older adults. The confirmed contract required return to Ya'an Railway Station and encoded medical, cultural, walking, activity, and rest requirements. Of 120 candidates, 102 passed the hard constraints, and the Senior-Traveler veto excluded six more. The remaining 96 formed the agreement set. The selected route was Pareto-efficient and maximized normalized gains for all three stakeholders, indicating low-conflict consensus (Supplementary Table S9). The route covered 5.22 km in 305.95 min, leaving 54.05 min within the budget. Additionally, QGIS loaded all 13 audit layers successfully. The deployed interface makes this audit chain visible to planners (Fig. 6). Together, these elements demonstrate computational and spatial traceability.

5 Discussion and conclusion

5.1 Discussion of findings

The results reveal four main findings about when LLM-supported itineraries can qualify as tourism action. First, factual grounding requires a separate authority boundary. Second, stakeholder claims are protected by screening unacceptable routes before final

ranking. Third, bargaining changes decisions mainly when stakeholder preferences diverge. Fourth, transparent abstention can open a bounded renegotiation of the request. Together, these findings recast itinerary generation as a question of action legitimacy: whether a route is evidentially grounded, acceptable to affected stakeholders, and eligible for implementation.

5.1.1 Factual grounding requires an authority boundary

Fig. 3 reveals a grounding–authority gap. The evidence-grounded LLM achieved 100% factual-claim accuracy and 45.14% protocol alignment, while recommending in all eight abstention scenarios. Under bounded authority, alignment reached 100% and forced recommendations fell to zero. Thus, legitimate tourism action combines factual reliability with bounded authority.

This pattern extends tourism research on misinformation, trust, and domain knowledge (Hsu et al. 2024; Kim et al. 2025b; Christensen et al. 2024; Kim et al. 2024; Foroughi et al. 2025; Wong et al. 2025). Grounding remains essential because route quality depends on verified opening times, travel distances, and service information. Action rules must then determine whether an inferred preference becomes binding, a protected threshold remains fixed, or the system issues a recommendation. These are allocations of decision rights. Consequently, responsible LLM design requires both an epistemic boundary around verified facts and an authority boundary around consequential choice.

Bounded generative agency preserves the useful flexibility of language models within this division. Older travelers and planners express needs in ordinary language, such as lower fatigue, cultural interest, or additional rest. The LLM can translate these expressions into a proposed contract and explain the computed outcome. Specifically, confirmation, GIS evidence, stakeholder rules, and the formal protocol determine the route action. The result refines the automation–augmentation debate: the central design question concerns both the degree of AI participation and the decision variables it is authorized to change (Raisch and Krakowski 2021; Shrestha et al. 2019; Baird and Maruping 2021; Fügener et al. 2022; Berente et al. 2021).

5.1.2 Stakeholder protection occurs before aggregation

Fig. 4a locates the main protection effect before ranking. Traditional methods selected stakeholder-ineligible routes in 17.5–20.0% of recommendable scenarios. Conversely, Nash, mean utility, and maximin retained 100% eligibility within the same agreement set. Removing the veto or acceptability gate reduced eligibility to 68.75% and 66.67%, respectively, and converted all eight abstentions into recommendations.

This finding explains why higher scalar utility alone can yield a weaker tourism decision. Organized senior itineraries distribute experience, delivery burdens, and destination responsibilities across several actors. Some claims are tradable, whereas health, mobility, schedule, or delivery minima may require separate standing. Aggregating first allows a gain for one actor to conceal an unacceptable loss for another, consistent with non-compensatory decision concerns (Bouyssou 1986; Roy 1991; Munda 2004). Acceptability-before-aggregation instead establishes which alternatives are eligible to become actions and then permits comparison within that set.

Accordingly, stakeholder-aware personalization extends beyond adding stakeholder weights to a common score. It keeps the welfare of the Senior Traveler, Destination Steward, and Tourism Operator visible until their minimum claims have been evaluated. The result moves personalization from preference fit toward implementable agreement. It also gives destination and operator interests a transparent role while retaining the older traveler as the primary beneficiary.

5.1.3 Bargaining becomes consequential as preferences diverge

Fig. 4b identifies a conflict boundary for bargaining. At $\gamma = 1$, Nash and mean utility selected the same route in all stress-test scenarios. At $\gamma = 3$, they differed in 15.0%, while Nash differed from Senior-Traveler-only and maximin in 32.5%. Thus, the results support a conditional, conflict-dependent role for Nash coordination.

This pattern clarifies the role of bargaining. The acceptability gates create stakeholder protection, while Nash structures proportional compromise after protected claims have been satisfied. Its smaller effect under aligned preferences therefore indicates existing consensus. Conversely, its growing decision impact under conflict identifies cases that require explicit coordination. For tourism managers, the relevant question is whether a planning case requires confirmation of common ground or active coordination of competing gains.

5.1.4 Abstention can open bounded renegotiation

The eight formal abstentions show that a completed tourism decision can appropriately yield responsible abstention. Across 52 time-boundary cases, action status changed with the available schedule. Additionally, all eight counterfactual repairs recovered an eligible route after a planner-confirmed change to time or stop count while preserving stakeholder protections (Fig. 5). Thus, abstention can reveal which confirmed requirement prevents action and what bounded adjustment would reopen the agreement set.

This finding turns abstention from a system endpoint into a governance interface. Predictive reject-option research ties withholding to decision reliability (Hendrickx et al. 2024; Hallberg Szabadváry et al. 2025); responsible abstention extends this logic to an empty protected agreement set. A transparent abstention exposes the distance between a request and an implementable route, allowing planners to revise scope deliberately while protected safety, mobility, and stakeholder thresholds remain fixed. Additionally, action status remained stable across the tested preference perturbations, while exact-route stability ranged from 84.01% to 91.74% (Supplementary Fig. S2). This distinction matters operationally: planners can treat recommendation eligibility as robust while still reviewing near-equivalent route variants.

5.2 Theoretical and methodological contributions

The findings connect representation, authority, and action into one account of LLM-enabled tourism planning. First, they extend personalization from matching the initiating user's preferences to preserving the claims of those who experience, deliver, and govern an implemented tourism product (Abdollahpouri et al. 2020; Sautter and

Leisen 1999; Font and Genzbigelytė-Venturi 2026). Multi-stakeholder personalization therefore depends on whose welfare remains visible and which claims retain standing.

Second, bounded generative agency identifies the decision right as a precise unit of generative-AI governance. Generic human oversight indicates that people remain involved, while the framework specifies whether an LLM may alter a confirmed constraint, stakeholder weight, candidate scope, or final action (Baird and Maruping 2021; Fügener et al. 2022; Berente et al. 2021). It assigns authority by decision stage. The grounding–authority gap further shows why this distinction is theoretically necessary: information quality and action authority govern separate functions.

Third, acceptability-before-aggregation extends tourism recommendation from relative ranking to action eligibility. The approach operationalizes legitimacy through a declared sequence: feasibility, stakeholder minima, agreement formation, coordination, and, where necessary, abstention. This sequence also yields a conditional account of bargaining. Non-compensatory gates protect stakeholder claims across cases, whereas Nash becomes more influential as conflict increases. Thus, legitimacy and compromise arise from different mechanisms.

Additionally, SilverRoute offers a methodological contribution by making these distinctions testable within an auditable LLM–GIS architecture. Its controlled experiments separately vary factual grounding, authority allocation, acceptability gates, and preference conflict, linking the theoretical claims to observable action outcomes. This design supports future tests of how authority maps vary with consequence, verifiability, and stakeholder claims.

5.3 Practical implications

For travel-agency product managers and organizations serving older adults, SilverRoute supports a clear division of work. The LLM can structure group needs and explain alternatives. Planners confirm the decision contract, while auditable procedures determine distance, feasibility, stakeholder acceptability, and final action. This division retains conversational accessibility while keeping formal action grounded in operational evidence.

For destination managers and tourism operators, decision traces and separate utilities reveal which stakeholder claim limits an otherwise attractive route. Route options should therefore be screened against non-negotiable minima before commercial ranking or bargaining. Additionally, preference-conflict scenarios can distinguish cases with shared priorities from those that require active compromise.

For tourism AI developers, evaluation should extend beyond factual accuracy. Protocol alignment, agreement eligibility, forced-recommendation rates, and the quality of abstention pathways are also decision-relevant. An authority registry can specify which fields the LLM may propose, which require confirmation, and which remain controlled by formal evidence and rules. When the agreement set is empty, the interface can report the blocking condition and offer minimum repairs for deliberate approval. Editable GIS outputs further connect these decisions to inspectable routes, facilities, and alternatives.

5.4 Conclusion and future research

SilverRoute reframes LLM itinerary planning as the governance of tourism action. Guided by bounded generative agency, it places generative interpretation and explanation around a confirmed decision contract and an auditable GIS-based multicriteria core. In the Agency Boundary Experiment, bounded authority achieved 100% protocol alignment and eliminated forced recommendations; additionally, all 40 issued recommendations remained agreement-eligible. The findings further show that legitimate action requires factual grounding plus bounded authority, stakeholder acceptability before ranking, conflict-sensitive bargaining, and explicit repairs after abstention. Accordingly, the framework can support travel-agency senior-tour product managers, local tourism and destination-management authorities, and organizations serving older adults in developing age-friendly cultural itineraries. It also keeps older travelers' safety, physical burden, and experience central to route design.

Future research can examine alternative bounded-authority maps across destinations, transport settings, and levels of decision consequence. Building on the controlled scenario analyses in Ya'an, comparative studies could assess how governance arrangements, service systems, and stakeholder priorities shape action eligibility. Additionally, process-oriented research could examine how organizational learning reshapes confirmed requirements and authority allocation over time.

Supplementary data

Online Resource 1 (PDF). Supplementary methods and results in Sections S1–S7, covering data preparation and provenance, stakeholder parameters, route settings, the validation registry, supplementary outcomes, the illustrative decision trace, and reproducibility. The associated computational package contains experiment summaries, figure source data, statistical tables, decision-transparency records, provenance metadata, and editable QGIS audit artifacts.

Data availability

Data supporting the findings are available from the corresponding author upon request.

References

Aaltonen S (2013) Trying to push things through: Forms and bounds of agency in transitions of school-age young people. Journal of Youth Studies 16(3):375–390. https://doi.org/10.1080/13676261.2012.733807

Abdollahpouri H, Adomavicius G, Burke R, et al (2020) Multistakeholder recommendation: Survey and research directions. User Modeling and User-Adapted Interaction 30(1):127–158. https://doi.org/10.1007/s11257-019-09256-1

Abou-Shouk M, Elbaz AM, Al-Leheabi SMZM, et al (2024) Trusting ChatGPT usage in personalized travel planning: The moderating role of privacy and data security. Tourism and Hospitality Research https://doi.org/10.1177/14673584241306346

Ali F, Petkovic Z, Lodhi RN, et al (2026) Age-friendly tourism as a dynamic service ecology: Older adults' lived experiences with travel, barriers, and adaptive strategies. International Journal of Tourism Research 28(4). https://doi.org/10.1002/jtr.70377

Amigó E, Deldjoo Y, Mizzaro S, et al (2023) A unifying and general account of fairness measurement in recommender systems. Information Processing & Management 60(1):103115. https://doi.org/10.1016/j.ipm.2022.103115

Baird A, Maruping LM (2021) The next generation of research on IS use: A theoretical framework of delegation to and from agentic IS artifacts. MIS Quarterly 45(1):315–341. https://doi.org/10.25300/MISQ/2021/15882

Banerjee A, Mahmudov T, Adler E, et al (2025) Modeling sustainable city trips: integrating CO_2e emissions, popularity, and seasonality into tourism recommender systems. Information Technology & Tourism 27:189–226. https://doi.org/10.1007/s40558-024-00303-1

Berente N, Gu B, Recker J, et al (2021) Managing artificial intelligence. MIS Quarterly 45(3):1433–1450. https://doi.org/10.25300/MISQ/2021/16274

Borràs J, Moreno A, Valls A (2014) Intelligent tourism recommender systems: A survey. Expert Systems with Applications 41(16):7370–7389. https://doi.org/10.1016/j.eswa.2014.06.007

Bouyssou D (1986) Some remarks on the notion of compensation in MCDM. European Journal of Operational Research 26(1):150–160. https://doi.org/10.1016/0377-2217(86)90167-0

Bramwell B, Lane B (2011) Critical research on the governance of tourism and sustainability. Journal of Sustainable Tourism 19(4-5):411–421. https://doi.org/10.1080/09669582.2011.580586

Chang CW, Tsai CY, Yao L, et al (2025) Personalized travel itinerary recommendation enhancing by user interests and point-of-interest characteristics. Information Technology & Tourism 27(3):649–682. https://doi.org/10.1007/s40558-025-00318-2

Chow CK (1970) On optimum recognition error and reject tradeoff. IEEE Transactions on Information Theory 16(1):41–46. https://doi.org/10.1109/TIT.1970.1054406

Christensen J, Hansen JM, Wilson P (2024) Understanding the role and impact of generative artificial intelligence (AI) hallucination within consumers' tourism decision-making processes. Current Issues in Tourism 28(4):545–560. https://doi.org/10.1080/13683500.2023.2300032

Cruz M, Jardim B, Neto MdC (2025) Lisa: a touristic chatbot for Lisbon. Information Technology & Tourism 27:1153–1183. https://doi.org/10.1007/s40558-025-00339-x

Dalla Vecchia A, Migliorini S, Quintarelli E, et al (2024) Promoting sustainable tourism by recommending sequences of attractions with deep reinforcement learning. Information Technology & Tourism 26:449–484. https://doi.org/10.1007/s40558-024-00288-x

Darcy S, Dickson TJ (2009) A whole-of-life approach to tourism: The case for accessible tourism experiences. Journal of Hospitality and Tourism Management 16(1):32–44. https://doi.org/10.1375/jhtm.16.1.32

Delic A, Neidhardt J, Nguyen TN, et al (2018) An observational user study for group recommender systems in the tourism domain. Information Technology & Tourism 19:87–116. https://doi.org/10.1007/s40558-018-0106-y

Diakoulaki D, Mavrotas G, Papayannakis L (1995) Determining objective weights in multiple criteria problems: The CRITIC method. Computers & Operations Research 22(7):763–770. https://doi.org/10.1016/0305-0548(94)00059-H

Dietz LW, Sánchez P, Bellogín A (2025) Understanding the influence of data characteristics on the performance of point-of-interest recommendation algorithms. Information Technology & Tourism 27:75–124. https://doi.org/10.1007/s40558-024-00304-0

Ding Y, Zhang L, Huang C, et al (2024) Two-stage travel itinerary recommendation optimization model considering stochastic traffic time. Expert Systems with Applications 237:121536. https://doi.org/10.1016/j.eswa.2023.121536

Dokoupil P, Boratto L, Peška L (2025) Accuracy and beyond-accuracy perspectives of controllable multi-objective recommender systems. Information Processing & Management 62(6):104267. https://doi.org/10.1016/j.ipm.2025.104267

Dwivedi YK, Pandey N, Currie W, et al (2024) Leveraging ChatGPT and other generative artificial intelligence (AI)-based applications in the hospitality and tourism industry: Practices, challenges and research agenda. International Journal of Contemporary Hospitality Management 36(1):1–12. https://doi.org/10.1108/IJCHM-05-2023-0686

Evans K (2007) Concepts of bounded agency in education, work, and the personal lives of young adults. International Journal of Psychology 42(2):85–93. https://doi.org/10.1080/00207590600991237

Fabris A, Rus C, Saldivar J, et al (2026) Does fair ranking lead to fair recruitment outcomes? a study of interventions, interfaces, and interactions. Information Processing & Management 63(3):104506. https://doi.org/10.1016/j.ipm.2025.104506

Fan Z, Chen C (2024) CuPe-KG: Cultural perspective-based knowledge graph construction of tourism resources via pretrained language models. Information Processing & Management 61(3):103646. https://doi.org/10.1016/j.ipm.2024.103646

Font X, Genzbigelytė-Venturi R (2026) How stakeholder co-creation reshaped EU tourism policy: The long-term legacy of the 2020 european tourism convention. Annals of Tourism Research 116:104113. https://doi.org/10.1016/j.annals.2025.104113

Foroughi B, Vu HTM, Thaichon P (2025) Building trust for sustained generative AI travel adoption. Tourism Review 81(2):718–754. https://doi.org/10.1108/TR-05-2025-0554

Fouad AM, Salem IE, Fathy EA (2024) Generative AI insights in tourism and hospitality: A comprehensive review and strategic research roadmap. Tourism and Hospitality Research 26(2):339–358. https://doi.org/10.1177/14673584241293125

Friedli J, Dolata M, Schwabe G (2025) How Heidi helps: designing for inspiration in digital travel assistants. Information Technology & Tourism 27:1047–1080. https://doi.org/10.1007/s40558-025-00328-0

Fügener A, Grahl J, Gupta A, et al (2022) Cognitive challenges in human–artificial intelligence collaboration: Investigating the path toward productive delegation. Information Systems Research 33(2):678–696. https://doi.org/10.1287/isre.2021.1079

Generosi A, Villafan JY, Ferretti M, et al (2025) A recommender-based web platform to boost tourism in marginal territories. Information Technology & Tourism 27:797–831. https://doi.org/10.1007/s40558-025-00327-1

Geninatti Cossatin A, Ferrero F, Ardissono L, et al (2026) The autonomy equation: How agentic AI reshapes trust and workload in routine productivity applications. Information Processing & Management 63(5):104681. https://doi.org/10.1016/j.ipm.2026.104681

Glikson E, Woolley AW (2020) Human trust in artificial intelligence: Review of empirical research. Academy of Management Annals 14(2):627–660. https://doi.org/10.5465/annals.2018.0057

Gretzel U, Fesenmaier DR (2006) Persuasion in recommender systems. International Journal of Electronic Commerce 11(2):81–100. https://doi.org/10.2753/JEC1086-4415110204

Hall CM (2011) A typology of governance and its implications for tourism policy analysis. Journal of Sustainable Tourism 19(4-5):437–457. https://doi.org/10.1080/09669582.2011.570346

Hallberg Szabadváry J, Löfström T, Johansson U, et al (2025) Classification with reject option: Distribution-free error guarantees via conformal prediction. Machine

Learning with Applications 20:100664. https://doi.org/10.1016/j.mlwa.2025.100664

Hendrickx K, Perini L, Van der Plas D, et al (2024) Machine learning with a reject option: a survey. Machine Learning 113(5):3073–3110. https://doi.org/10.1007/s10994-024-06534-x

Hsu CHC, Tan G, Stantic B (2024) A fine-tuned tourism-specific generative AI concept. Annals of Tourism Research 104:103723. https://doi.org/10.1016/j.annals.2023.103723

Huang GI, Zhang CJ, Zhou Torres WC, et al (2026) Responsible AI and human collaboration in tourism management: Ethical considerations and identity disclosure. Tourism Management 113:105315. https://doi.org/10.1016/j.tourman.2025.105315

Hwang CL, Yoon K (1981) Multiple Attribute Decision Making: Methods and Applications—A State-of-the-Art Survey. Springer, https://doi.org/10.1007/978-3-642-48318-9

Jankowski P (1995) Integrating geographical information systems and multiple criteria decision-making methods. International Journal of Geographical Information Systems 9(3):251–273. https://doi.org/10.1080/02693799508902036

Jeng J, Fesenmaier DR (2002) Conceptualizing the travel decision-making hierarchy: A review of recent developments. Tourism Analysis 7(1):15–32. https://doi.org/10.3727/108354202108749925

Jia S, Ma C, Chi OH, et al (2026) A socio-technical exploration of bias in generative AI travel planners. Tourism Management 116:105442. https://doi.org/10.1016/j.tourman.2026.105442

Keeney RL, Raiffa H (1976) Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Wiley

Khaitan S, Shrivastava R (2026) Developing fairness, accuracy, and serendipity objective functions for recommendation system and establishing trade-off through multi-objective evolutionary optimization. Information Processing & Management 63(4):104604. https://doi.org/10.1016/j.ipm.2025.104604

Kim JH, Kim J, Baek TH, et al (2025a) ChatGPT personalized and humorous recommendations. Annals of Tourism Research 110:103857. https://doi.org/10.1016/j.annals.2024.103857

Kim JH, Kim J, Park J, et al (2025b) When ChatGPT gives incorrect answers: The impact of inaccurate information by generative AI on tourism decision-making. Journal of Travel Research 64(1):51–73. https://doi.org/10.1177/00472875231212996

Kim MJ, Kang SE, Hall CM, et al (2024) Unveiling the impact of ChatGPT on travel consumer behaviour: exploring trust, attribute, and sustainable-tourism action. Current Issues in Tourism 28(8):1191–1196. https://doi.org/10.1080/13683500.2024.2355556

Leung XY, Zhong YS, Sun J (2025) The impact of social media influencer's age cue on older adults' travel intention: The moderating roles of travel cues and travel constraints. Tourism Management 106:104979. https://doi.org/10.1016/j.tourman.2024.104979

Li TE, Chan ETH, Wood E (2024) Navigating challenges: coping, memory, and the interpretations of travel adversities among middle-aged and older chinese adults. Current Issues in Tourism 29(5):867–884. https://doi.org/10.1080/13683500.2024.2440616

Lu X, Su X, Feng J, et al (2026) A framework for hallucination mitigation in domain-specialized large language models with application to aviation maintenance decision support. Information Processing & Management 63(6):104762. https://doi.org/10.1016/j.ipm.2026.104762

Malczewski J (2006) GIS-based multicriteria decision analysis: a survey of the literature. International Journal of Geographical Information Science 20(7):703–726. https://doi.org/10.1080/13658810600661508

Mikulić J (2025) Large language models in tourism research: Support, substitution, or something else? Annals of Tourism Research 114:104003. https://doi.org/10.1016/j.annals.2025.104003

Munda G (2004) Social multi-criteria evaluation: Methodological foundations and operational consequences. European Journal of Operational Research 158(3):662–677. https://doi.org/10.1016/S0377-2217(03)00369-2

Nash JF (1950) The bargaining problem. Econometrica 18(2):155–162. https://doi.org/10.2307/1907266

Nicolau JL, Losada N, Alén E, et al (2020) The staged nature of decision making among senior tourists. Journal of Travel Research 59(4):602–613. https://doi.org/10.1177/0047287519851229

Opricovic S, Tzeng GH (2004) Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. European Journal of Operational Research 156(2):445–455. https://doi.org/10.1016/S0377-2217(03)00020-1

Otaki K, Baba Y (2025) Travel itinerary recommendation using interaction-based augmented data. Expert Systems with Applications 269:126294. https://doi.org/10.1016/j.eswa.2024.126294

Pan L, Ge Z, Ma J (2025) Understanding travel constraints, negotiation strategy and travel expenditure: An empirical study of chinese senior tourism. Tourism Management Perspectives 58:101397. https://doi.org/10.1016/j.tmp.2025.101397

Park SY, Fesenmaier DR (2014) Travel decision flexibility. Tourism Analysis 19(1):35–49. https://doi.org/10.3727/108354214X13927625340154

Raisch S, Krakowski S (2021) Artificial intelligence and management: The automation–augmentation paradox. Academy of Management Review 46(1):192–210. https://doi.org/10.5465/amr.2018.0072

Ramrakhiyani N, Varma V, Palshikar GK, et al (2025) Gauging, enriching and applying geography knowledge in pre-trained language models. Information Processing & Management 62(1):103892. https://doi.org/10.1016/j.ipm.2024.103892

Roy B (1991) The outranking approach and the foundations of ELECTRE methods. Theory and Decision 31(1):49–73. https://doi.org/10.1007/BF00134132

Satyanarayan A, Jones GM (2024) Intelligence as agency: Evaluating the capacity of generative AI to empower or constrain human action. An MIT Exploration of Generative AI https://doi.org/10.21428/e4baedd9.2d7598a2

Sautter ET, Leisen B (1999) Managing stakeholders: A tourism planning model. Annals of Tourism Research 26(2):312–328. https://doi.org/10.1016/S0160-7383(98)00097-8

Shrestha YR, Ben-Menahem SM, von Krogh G (2019) Organizational decision-making structures in the age of artificial intelligence. California Management Review 61(4):66–83. https://doi.org/10.1177/0008125619862257

Tang R, Qu Z, Zhao S, et al (2025) Sustainable service of young tour guides for senior tourists: An exploration based on intergenerational relationships and intergroup contact. Journal of Travel Research https://doi.org/10.1177/00472875251387262

Tao CB, Huang SS, Wang J, et al (2024) Accessible tourism: tourists with physical disability—segmentation based on perceived travel barriers. Tourism Review 80(9):1654–1676. https://doi.org/10.1108/TR-07-2023-0459

Trilaksono T, Dewantara MH, Liu B, et al (2025) Inclusive business practice in local tourism ecosystems: stakeholder collaboration and rebranding commercial firms. Tourism Review https://doi.org/10.1108/TR-01-2025-0048

Tussyadiah I (2020) A review of research into automation in tourism: Launching the annals of tourism research curated collection on artificial intelligence and robotics in tourism. Annals of Tourism Research 81:102883. https://doi.org/10.1016/j.annals.2020.102883

Wang P, Ding H (2024) The rationality of explanation or human capacity? understanding the impact of explainable artificial intelligence on human–ai trust and decision performance. Information Processing & Management 61(4):103732. https://doi.org/10.1016/j.ipm.2024.103732

Wang Y, Miao H, Xiong M, et al (2025) How does ChatGPT-generated information influence travel planning? the mediating role of inspiration. Tourism Review 81(2):694–717. https://doi.org/10.1108/TR-03-2025-0235

Wong IA, Lian QL, Sun D (2023) Autonomous travel decision-making: An early glimpse into ChatGPT and generative AI. Journal of Hospitality and Tourism Management 56:253–263. https://doi.org/10.1016/j.jhtm.2023.06.022

Wong JWC, Lai IKW, Lin Y (2025) The perceived reliability and adoption intention towards human-generated content vs. AI-generated content for travel planning: a moderating role of travel persona. Journal of Travel & Tourism Marketing 42(4):461–478. https://doi.org/10.1080/10548408.2025.2468468

Xu B, Li X, Fan ZP (2024) Selection and visiting sequence of daily attractions: Multi-day travel itinerary recommendation based on multi-source online data. Expert Systems with Applications 250:123895. https://doi.org/10.1016/j.eswa.2024.123895

Yao J, Pang Q, Zhang Y (2025) Leveraging digital technology for integration: innovating tourism business models for multi-stakeholder value co-creation. Current Issues in Tourism pp 1–19. https://doi.org/10.1080/13683500.2025.2590590

Zhao H, Yuan B, Zhang B, et al (2026) Navigating generative AI use in tourism and hospitality: how trust and ethics shape human-machine information interaction of practitioners. Tourism Review pp 1–23. https://doi.org/10.1108/TR-07-2025-0740

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1.pdf