Journal of Consumer Behaviour

WILEY

RESEARCH ARTICLE OPEN ACCESS

AI Hallucinations in Tourism: How Errors Impact Consumer Trust and Recommendation Acceptance

Francisco Rejón-Guardia[1,2] | Sebastian Molinillo[1,2] | Rafael Anaya-Sánchez[1,2]

[1]Andalusian Institute for Research and Innovation in Tourism of the Universities of Granada, Malaga and Seville, Malaga, Spain | [2]Institute of Economics and Business for Society, University of Malaga, Malaga, Spain

Correspondence: Francisco Rejón-Guardia (franrejon@uma.es)

Received: 30 April 2025 | **Revised:** 12 November 2025 | **Accepted:** 23 December 2025

Keywords: AI hallucinations | behavioural intentions | ChatGPT | GenAI | generative artificial intelligence | perceived accuracy | tourism planning | user trust

ABSTRACT

Generative artificial intelligence (GenAI) is quickly transforming travel planning; however, its outputs can include hallucinations, which are plausible yet false statements that can undermine user judgement. Eliminating hallucinations in GenAI technology is currently impossible. This research analyses, through two experiments, how hallucinations in itineraries generated by ChatGPT influence consumer behaviour. Study 1 ($n = 1004$) explores a serial-mediation pathway where the presence of hallucinations affects the intention to follow the itinerary through perceived accuracy, perceived usefulness and trustworthiness. Hallucinations significantly reduced perceived accuracy. As accuracy increased, usefulness improved, leading to greater trustworthiness, which strongly predicted intention. Although the direct effect of hallucinations on intention was not significant, the serial indirect effect was negative and significant. Study 2 ($n = 241$) investigates how the outcome's positive or negative value and the importance of error feedback influence the results. The findings indicate that the valence of outcomes moderates the relationship (i.e., positive outcomes yielded higher intention to follow the recommendation than negative outcomes). However, when errors were salient (i.e., when there was a hallucination accompanied by negative consequences) intentions to follow the recommendation decreased further. Together, these studies contribute to an integrative framework that connects cognitive factors (accuracy, usefulness), attitudinal factors (trust), experiential factors (observed outcomes) and contextual factors (error salience) to explain how and when GenAI hallucinations can reduce compliance with travel advice. The findings enhance our understanding of the potential drawbacks of GenAI in tourism and provide actionable guidance for developing more transparent, reliable and user-centred GenAI travel systems.

1 | Introduction

Generative artificial intelligence (GenAI) has emerged as a disruptive technology profoundly transforming how consumers plan their trips and access tourism-related information. Tools such as ChatGPT or Gemini, built upon large-scale language models (LLMs), enable the generation of personalised itineraries, tailored recommendations and detailed destination descriptions, thereby enhancing planning efficiency and raising expectations regarding the quality of suggestions received (Christensen et al. 2024; Kim, Kim, Kim, and Kim 2023; Kim, Kim, Kim, and Hailu 2023; Yaprak 2024).

Specifically, the adoption of GenAI-based technologies within the tourism sector is projected to grow at an annual rate of 30% until 2030, fuelled by the demand for personalised, efficient and real-time services (Yaprak 2024). Unlike traditional sources of tourism information, such as official websites, opinion aggregators or human recommendations, GenAI systems stand out due to their ability to simulate natural

language, incorporate user preferences and deliver instant responses, thereby enhancing perceived personalisation and utility (Christensen et al. 2024; Wong et al. 2023). This revolution in human–technology interaction has triggered growing academic interest in the behavioural effects of using these tools in tourism contexts (Jia et al. 2024).

Whilst the potential benefits of GenAI in tourism planning are considerable, scholars have also highlighted its dark side. Beyond efficiency gains, GenAI systems introduce risks linked to misinformation, lack of transparency and the erosion of consumer autonomy (Grewal et al. 2021; Puntoni et al. 2021). These concerns are particularly acute in tourism, where decisions often depend on the accuracy of recommendations. The very features that make GenAI valuable to consumers also generate unintended costs. According to Grewal et al. (2021), AI can compromise transparency and fairness, whilst Puntoni et al. (2021) identify hidden costs such as loss of control, exploitation and alienation. Together, these insights frame GenAI not as a neutral tool but as an experiential technology with profound social and psychological consequences.

One of the most pressing risks is the phenomenon of GenAI 'hallucinations', defined as the generation of false, fabricated or misleading information by a model, often presented with high confidence (Sun et al. 2024). These outputs may range from minor factual inaccuracies to elaborate yet fictitious content, creating serious challenges in domains such as travel planning. In tourism, where consumers frequently rely on accurate information to decide itineraries, accommodation or activities, even seemingly minor errors can undermine trust and damage user experiences (Christensen et al. 2024; Kim, Kim, Kim, and Kim 2023; Kim, Kim, Kim, and Hailu 2023). For example, Christensen et al. (2024) show that consumers sometimes adopt itineraries containing obvious GenAI-generated errors, perceiving them as neutral or personalised, which reveals a paradox of over-reliance on automated systems. Common inaccuracies include fictitious attraction opening hours or references to nonexistent restaurants, which can generate frustration, financial loss and reputational damage for destinations or brands (Kim et al. 2025; Yaprak 2024). In this respect, hallucinations not only compromise the credibility of the technology but also pose broader risks to consumer well-being.

Although some studies demonstrate that GenAI errors reduce visitation intentions (Kim et al. 2025), the underlying mechanisms remain underexplored. Hallucinations can mislead users, disrupt trust and impair decision quality (Christensen et al. 2024; Sun et al. 2024). These risks are magnified by the opaque, black-box nature of GenAI systems, which restrict user control and transparency (Grewal et al. 2021). As Dietvorst et al. (2015) note, such shortcomings may trigger algorithm aversion, whilst Puntoni et al. (2021) stress that they can also erode consumer autonomy and emotional stability. Consequently, hallucinations represent a critical expression of the dark side of AI in tourism contexts.

Despite the growing deployment of GenAI for travel planning, important gaps persist. Kim et al. (2025) demonstrate that errors can dampen destination visit intentions; however, we lack clarity on how specific error characteristics and their real-world consequences influence consumer responses. In particular, little is known about how travellers interpret and react to different error types (e.g., fabrications such as a non-existent restaurant vs. factual inaccuracies such as incorrect opening hours), how error salience matters (e.g., high when a closed venue leads to a wasted trip vs. low when a minor date mismatch is inconsequential) and what downstream outcomes follow (e.g., cancelled bookings, time loss, extra expense or even serendipitous discovery of an enjoyable alternative). Accordingly, we formulate the following research questions: How do error type and error salience differentially affect trust, perceived reliability and intentions to use or book? Through which psychological pathways do hallucinations influence judgements and choices? Moreover, do tangible outcomes (wasted trip, time loss, extra expense) mediate or moderate the relationship between hallucinations and consumer behaviour? The present study addresses this gap by examining the cognitive, attitudinal, behavioural and experiential consequences of hallucination-induced errors in tourism recommendations, using two controlled experiments.

This research makes three significant contributions. First, it builds on previous studies by examining not only the occurrence of hallucinations but also the effects of experiencing their positive outcomes (unexpectedly satisfactory experiences) and negative outcomes (dissatisfaction and unmet expectations). Second, it moves beyond traditional structural views on errors (such as type and relevance) by analysing how the valence of the outcome and the salience of the error together influence consumer acceptance. Third, it develops a comprehensive framework that articulates cognitive, attitudinal, contextual and other dimensions, providing a deeper understanding of the darker aspects of GenAI in tourism and expanding beyond standard acceptance models.

The article is structured as follows: first, it presents a theoretical review of GenAI and its implications for digital tourism. Next, it outlines the experimental design and the variables being investigated. Then, it examines the empirical results obtained. Finally, it discusses the theoretical and practical implications of the study, as well as its limitations and potential directions for future research.

2 | Theoretical Background and Hypotheses Development

2.1 | Conceptualisation and Causes

GenAI has revolutionised travel planning by enabling the real-time generation of personalised recommendations through models and services such as ChatGPT, Gemini or DeepSeek. These tools surpass traditional sources, such as official websites or forums, by offering adaptive itineraries, simulated conversations and immediate solutions (Christensen et al. 2024; Kim, Kim, Kim, and Kim 2023; Kim, Kim, Kim, and Hailu 2023). Their capacity to tailor content to individual interests positions GenAI as a key catalyst for innovation in the tourism experience (Yaprak 2024). Beyond this, GenAI tools can support tourists across all stages of travel, from the pre-trip planning phase, where they are particularly valuable for information search and itinerary design, to in-trip

assistance and post-trip reflection. During the early stages, GenAI reduces information overload by synthesising vast amounts of data into coherent, personalised recommendations (Carvalho and Ivanov 2023; Liu and Du 2025). In-trip, its ubiquity enables travellers to receive real-time guidance and problem-solving, enhancing perceived control and reducing uncertainty (Solomovich and Abraham 2024; Xu et al. 2025). Post-trip, its interactive and anthropomorphic qualities foster engagement and emotional connection, which can influence satisfaction and word-of-mouth behaviour (Gursoy et al. 2023; Wong et al. 2023). This multi-stage utility underscores the transformative potential of GenAI in shaping the end-to-end tourism experience.

However, the widespread deployment of these technologies has also exposed new vulnerabilities, particularly concerning the accuracy of the information produced. Unlike human sources, GenAI models can generate unintentional errors, posing specific challenges in contexts where accuracy and reliability are critical for decision-making, such as travel planning (Guttentag et al. 2024; Hrankai and Mak 2025; Ji et al. 2023; van Dis et al. 2023). Specifically, one of the primary risks associated with GenAI is the production of hallucinations, that is, coherent yet incorrect, fabricated or logically inconsistent outputs presented with undue confidence (Ji et al. 2023; Sun et al. 2024). Hallucinations come from three main causes. First, problems in the training data, such as mistakes or biases, lead models to repeat false patterns or unverified information (Ji et al. 2023; Rawte et al. 2023). Second, models tend to be overconfident and are designed to prioritise fluent, natural-sounding text over factual accuracy, which makes them produce confident but incorrect answers (Kalai and Vempala 2024; OpenAI 2024). Third, because these systems do not reason deeply or connect their answers to real sources, they often rely on stored patterns in their parameters instead of evidence, especially on tasks requiring causal explanations or 'what if' reasoning (Ji et al. 2023; van Dis et al. 2023). Notably, both recent theoretical analyses (Kalai and Vempala 2024) and industry reports highlight that hallucinations are not merely residual flaws but an inherent by-product of how LLMs are trained and evaluated. As OpenAI (2025) explicitly recognises, hallucinations remain a fundamental challenge that cannot be entirely eliminated, since current training and evaluation methods incentivise confident guessing over calibrated uncertainty. Thus, efforts can only aim to mitigate rather than fully eradicate hallucinations, underscoring their persistent role in GenAI systems.

In tourism, these mechanisms yield factual inaccuracies, fabrications, omissions or illogical chains (e.g., non-existent venues, wrong opening hours), which degrade decision quality and trust (Christensen et al. 2024; Kim et al. 2025; Sun et al. 2024). Notably, some users still adopt GenAI suggestions due to perceived neutrality or personalisation, masking underlying risk (Christensen et al. 2024). GenAI hallucinations and classic recommendation errors differ in nature, impact and mechanism. Hallucinations concern factual veracity: the system asserts plausible but false or ungrounded information (intrinsic: contradicts a source; extrinsic: unverifiable by the source) (Ji et al. 2023; Sun et al. 2024). By contrast, traditional recommender errors typically involve relevance, coverage or ranking, suggesting options that are suboptimal or poorly matched, without inventing facts (Xu et al. 2024).

The trust consequences diverge accordingly: hallucinations directly erode credibility because they violate the truthfulness expectation of an informational agent, whereas classic recommender misfires more often produce dissatisfaction or choice overload rather than a breach of epistemic trust (Christensen et al. 2024; Kim et al. 2025; Xu et al. 2024). Systematically, hallucinations stem from LLMs' fluency-first decoding and weak grounding, whereas classical recommendation errors reflect data sparsity, preference misestimation or ranking noise (Ji et al. 2023; Sun et al. 2024). In tourism, this distinction matters because factually wrong content misleads itineraries and damages destinations, amplifying negative behavioural responses (Christensen et al. 2024; Kim et al. 2025) (see Appendix A).

In the tourism domain, hallucinations can result in user frustration, resource loss and distrust (Kumar et al. 2023). Nevertheless, Christensen et al. (2024) observed that even when users identify hallucinations, they may continue to use GenAI due to its perceived neutrality. Other studies, however, indicate that the presence of hallucinations significantly diminishes users' intention to follow recommendations (Kim et al. 2025) and may elicit negative emotional responses (Yaprak 2024).

The accessibility-diagnosticity theory, proposed by Feldman and Lynch Jr. (1988), posits that the influence of information on decision-making depends on two key factors: its accessibility, how easily it can be retrieved, and its diagnosticity, its perceived relevance to the task at hand. This framework is particularly useful in understanding the impact of hallucinations generated by GenAI models such as ChatGPT in tourism decision contexts (Kim, Kim, Kim, and Kim 2023; Kim et al. 2025). When erroneous GenAI-provided information is both easily accessible and perceived as diagnostic, it is more likely to distort user judgements, undermining not only specific decisions but also overall trust in the system. In this regard, the salience of the error, defined by its visibility, clarity or perceived impact, alongside users' interpretation of the error as indicative of overall system performance, is a critical factor amplifying the detrimental effects of hallucinations. Consequently, the severity of the impact is not solely determined by the error itself but also by its perceived relevance to the user's task.

Thus, the accessibility-diagnosticity theory gains salience in the domain of GenAI tourism recommendations, where information accuracy and credibility are rigorously scrutinised, especially under conditions of perceived uncertainty or risk (McKnight et al. 2011; Shi et al. 2021). Some scholars argue that, unlike human sources, errors committed by GenAI systems may provoke a sharper decline in trust due to the so-called algorithm aversion effect (Dietvorst et al. 2015). Consistent with this, recent studies show that the presence of GenAI errors significantly reduces users' willingness to follow recommendations, even when users acknowledge the error, as it is perceived as a systemic flaw that undermines trust in the technology overall (Christensen et al. 2024; Kim

et al. 2025; Sun et al. 2024) (see Appendix A). Accordingly, the following hypothesis is proposed:

H1. *The presence of hallucinations in ChatGPT-generated recommendations decreases users' likelihood of following the suggested itinerary.*

2.2 | Cognitive Mediators: Perceived Accuracy, Usefulness and Trust

Prior research on technology adoption consistently shows that cognitive evaluations are the primary engine behind users' behavioural intentions (e.g., perceived accuracy, usefulness and trust) (Jia et al. 2024; Kim et al. 2025; Shi et al. 2021; Topsakal 2025). In information-provision contexts, perceived accuracy acts as a first-stage filter through which people assess credibility and decide whether to act (Filieri and McLeay 2014; McKnight et al. 2011; Park et al. 2007). Classic acceptance frameworks, the Technology Acceptance Model (TAM) and the Theory of Planned Behaviour (TPB), likewise position belief formation as antecedent to intention, with judgements about information quality shaping downstream attitudes and behaviour (Ajzen 1985; Davis 1989). Applied to conversational recommenders such as ChatGPT, this logic implies that any factor that degrades perceived accuracy should also erode subsequent evaluations and intentions.

GenAI hallucinations are precisely such a factor: they are plausible yet false, unfounded or logically inconsistent outputs often delivered with unwarranted confidence (Christensen et al. 2024; Kumar et al. 2023; OpenAI 2024; Sun et al. 2024; Yaprak 2024). Hallucinations arise from data noise and bias, fluency-over-truth optimisation and shallow, ungrounded generation, making them difficult for lay users to detect (Ji et al. 2023; OpenAI 2024; Sun et al. 2024). In travel settings, where informational precision is integral to plans and costs, exposure to fabricated venues, incorrect hours or misleading routes directly undermines accuracy judgements and, in turn, perceived reliability and willingness to act on the advice (Christensen et al. 2024; Hrankai and Mak 2025; Kim, Kim, Kim, and Kim 2023; Kim, Kim, Kim, and Hailu 2023; Kim et al. 2025). Based on this literature, we propose the following hypothesis:

H2. *The presence of hallucinations negatively affects the perceived accuracy of ChatGPT-generated recommendations.*

Prior research consistently highlights the central role of cognitive evaluations in shaping users' behavioural intentions towards emerging technologies (Jia et al. 2024; Kim et al. 2025; Shi et al. 2021; Topsakal 2025). Within this framework, perceived accuracy functions as the primary filter through which users assess the credibility and value of information (Hrankai and Mak 2025; Kim, Kim, Kim, and Kim 2023; Kim et al. 2025). Consumers tend to regard objectively correct information as more valuable for decision-making because it enhances precision and efficiency, ultimately saving time and effort (Demir and Demir 2023; Guttentag et al. 2024; Park et al. 2007; Wong et al. 2023). Indeed, prior studies demonstrate that users perceive GenAI assistants as both intelligent and useful when they consistently provide accurate and high-quality recommendations (Hrankai and Mak 2025).

By contrast, when GenAI recommendations are inaccurate or contain hallucinations, users are less likely to perceive them as helpful for itinerary planning (Kim, Kim, Kim, and Kim 2023; Loureiro et al. 2024; Shin et al. 2025). The presence of incorrect information undermines both accuracy and reliability perceptions, thereby reducing consumers' willingness to adopt or follow GenAI-generated recommendations (Kim et al. 2025; Pretolesi et al. 2024). In this regard, prior research has shown that the accuracy of ChatGPT-generated content significantly influences perceived usefulness and, consequently, acceptance of recommendations, positioning accuracy as a key antecedent of technology adoption (Kim, Kim, Kim, and Kim 2023; Kim, Kim, Kim, and Hailu 2023; Topsakal 2025). Accuracy judgements subsequently shape perceptions of usefulness, as users tend to consider information that is factually correct as more valuable for decision-making (Park et al. 2007). In contrast, when recommendations are inaccurate, users are less likely to perceive them as helpful for planning their itineraries. Accordingly, we propose the following hypothesis:

H3. *Perceived accuracy positively influences the perceived usefulness of ChatGPT-generated recommendations.*

Trust emerges as a critical attitudinal construct shaping the acceptance and sustained use of GenAI technologies (Christensen et al. 2024; Duong et al. 2025; Kim, Kim, Kim, and Kim 2023; Kim et al. 2025; Liu and Du 2025; Shi et al. 2021; Topsakal 2025). Within this framework, information perceived as useful strengthens users' trust, as they interpret the technology as capable of supporting their goals, reducing uncertainty, and facilitating effective decision-making (Chang and Park 2024; Demir and Demir 2023; Hrankai and Mak 2025; Wong et al. 2023). Empirical evidence reinforces this connection, demonstrating that the quality, accuracy and relevance of GenAI recommendations are significant antecedents of user trust. For instance, the perceived relevance, credibility and usefulness of ChatGPT outputs have been shown to positively influence behavioural intentions through enhanced trust (Ali et al. 2023; Guttentag et al. 2024). Similarly, perceived system performance has been identified as a strong predictor of trust in GenAI recommenders (Chang and Park 2024). When users believe that GenAI technologies consistently meet their needs and expectations, they develop stronger identification with the system, thereby reinforcing trust (Liu and Du 2025). Conversely, exposure to hallucinations undermines perceived system reliability, reducing users' trust and highlighting the need for consistent accuracy to maintain confidence (Kim, Kim, Kim, and Kim 2023; Kim et al. 2025; Loureiro et al. 2024; Pretolesi et al. 2024; Shin et al. 2025). In tourism contexts, where decision-making is inherently uncertain, ChatGPT's ability to deliver personalised, timely and precise recommendations is essential for generating value and fostering traveller trust (Demir and Demir 2023; Gursoy et al. 2023). Accordingly, we propose the following hypothesis:

H4. *Perceived usefulness positively influences users' trustworthiness in ChatGPT.*

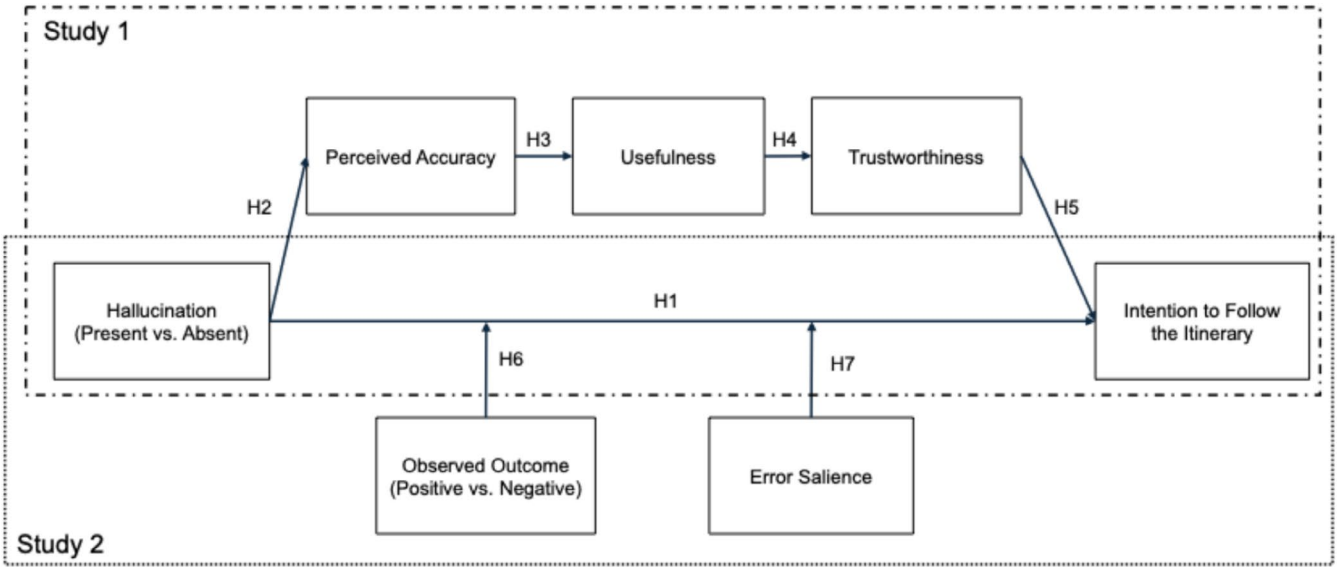

FIGURE 1 | Theoretical framework.

Incorrect or fabricated information produced by GenAI is perceived as less precise and less trustworthy, negatively influencing travellers' intention to visit suggested destinations (Kim et al. 2025; Loureiro et al. 2024). Users are less inclined to trust or act upon recommendations if they perceive the GenAI as incompetent or lacking common sense (Hrankai and Mak 2025). Thus, the ability of ChatGPT to consistently provide precise and useful content is fundamental to maintaining the credibility of its recommendations and ensuring that users regard them as reliable. Accordingly, we propose the following hypothesis:

H5. *Trustworthiness positively influences intention to follow the itinerary.*

2.3 | Experiential Moderators: Observed Outcomes and Error Salience

In addition to the informational content itself, the user's experience after following a GenAI-generated recommendation significantly influences their subsequent evaluation of the system. Multiple studies have shown that the observed positive or negative outcomes act as experiential anchors that moderate the perception of prior errors (e.g., a positive outcome could be discovering an enjoyable attraction despite an inaccurate recommendation, whereas a negative outcome could be arriving at a non-existent restaurant) (Kim et al. 2025). This effect aligns with experiential learning theories and post-use trust models, which argue that specific experiences with a technology shape future willingness to use or trust it (Filieri and McLeay 2014; McKnight et al. 2002; Shi et al. 2021). Thus, when an erroneous recommendation leads to a satisfactory experience, users tend to downplay the mistake and maintain their trust in the system (e.g., a wrong opening hour is ignored if the restaurant is still accessible and enjoyable). Conversely, suppose the error results in a negative experience (i.e., with the error being salient). In that case, it is interpreted as a systemic flaw, amplifying its impact and deteriorating perceived credibility (e.g., arriving at a recommended hotel that does not exist) (Christensen et al. 2024; Kim et al. 2025; Sun et al. 2024). This phenomenon highlights the role of practical experience in the acceptance of generative GenAI-based systems and in shaping user behaviour regarding the information they provide. Based on this rationale, the following hypothesis is proposed:

H6. *The valence of the observed consequence after following a ChatGPT-generated recommendation moderates the effect of errors, such that positive outcomes mitigate their negative impact on the intention to follow the suggested itinerary, whereas negative outcomes amplify this effect.*

Information that is highly accessible and relevant in memory strongly influences evaluations and behaviours. In this context, error salience, defined as the combination of incorrect information generated by ChatGPT (hallucinations) and the subsequent observation of a negative outcome (e.g., a recommended restaurant that turns out to be closed or non-existent), increases the accessibility and diagnosticity of the failure, thereby intensifying its impact on consumers' intention to follow the recommendation (Feldman and Lynch Jr. 1988). This relationship is consistent with previous research showing that when errors in GenAI content are detected, especially when they lead to unfavourable consequences (e.g., wasted time, cancelled bookings or financial loss), users' trust, perceived credibility and intention to reuse the system are substantially reduced (Christensen et al. 2024; Kim et al. 2025; Yaprak 2024). Moreover, the literature on algorithm aversion indicates that consumers evaluate mistakes made by automated systems more harshly than those made by humans, particularly when such errors cause negative outcomes (Dietvorst et al. 2015). Therefore, the following hypothesis is proposed:

H7. *Error salience moderates the effect of incorrect information, such that when a hallucination leads to a negative outcome, consumers show a significant reduction in their intention to follow the suggested itinerary.*

The overall study framework is depicted in Figure 1.

3 | Methodology

3.1 | Research Overview

We conducted two complementary experiments: Study 1 ($n = 1004$; September 2024) investigates the mediational pathway from the presence of hallucinations to the intention to follow the itinerary, focusing on perceived accuracy, perceived usefulness and trustworthiness. Study 2 (a between-subjects design with a 2×2 configuration: hallucination present vs. absent × observed outcome positive vs. negative; $n = 241$; July 2025) explores the moderating effects of outcome valence (either a positive or negative consequence following the GenAI recommendation) and error salience (the combination of hallucinations and negative outcomes that makes the error highly visible or diagnostic to participants). Both studies utilised real screenshots from ChatGPT as stimuli (refer Figures 2 and 3 for details).

ChatGPT functions as a personalised recommendation agent capable of suggesting routes, activities and travel options aligned with individual preferences (Kim, Kim, Kim, and Kim 2023; Wong et al. 2023; Yaprak 2024). By combining advanced natural language processing with vast training corpora, ChatGPT elevates tourism digitalisation to a new level, offering hyper-personalised planning experiences that rival and, in some respects, surpass those provided by human or institutional sources (Christensen et al. 2024).

3.2 | Study 1: Analysing the Mediational Pathway From the Presence of Hallucinations to the Intention to Follow the Itinerary: Perceived Accuracy, Perceived Usefulness and Trustworthiness

In Study 1, we aim to establish the baseline causal effect of GenAI hallucinations on consumers' acceptance of ChatGPT's itinerary. We expect that the presence of hallucinations will reduce participants' intention to follow the recommended itinerary because errors lower perceived accuracy, diminish perceived usefulness and erode trust in the recommendation. This Study 1 adopted a between-subjects design with two conditions (hallucination: present vs. absent).

3.2.1 | Participants, Procedure and Measurement

We conducted an a priori power analysis using G*Power (Kang 2021) to estimate the minimum sample size needed to detect the targeted effect with adequate power. We employed an F test for linear multiple regression (fixed model, examining R^2 deviation from zero) with an assumed large effect size ($f^2 = 0.35$), an alpha level of 0.01 and a desired statistical power of 0.95. The analysis indicated that a minimum sample size of 77 participants was needed to achieve the specified power. The sample size exceeded this threshold, ensuring adequate power to detect both main and mediation effects.

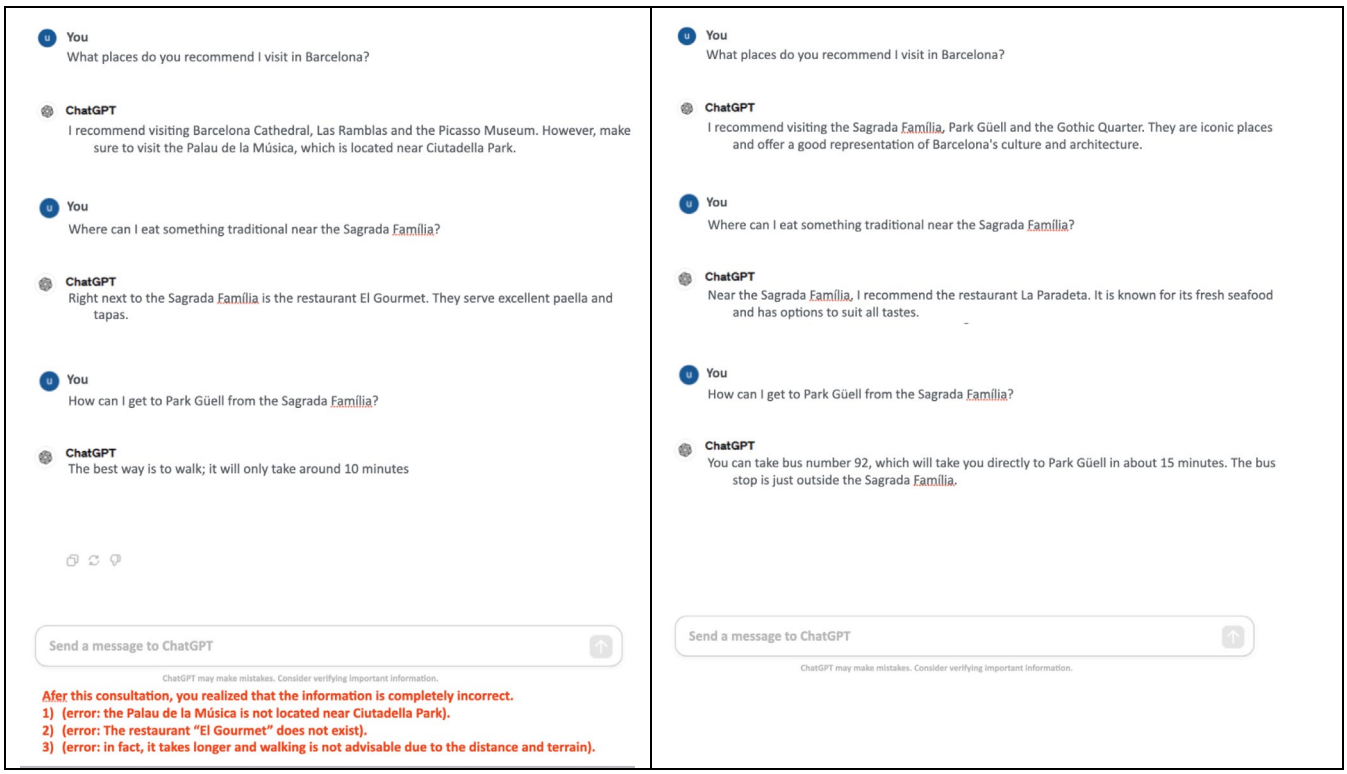

FIGURE 2 | Stimuli for Study 1. Left stimulus with incorrect information (hallucinations present). Right stimulus without incorrect information (no hallucinations).

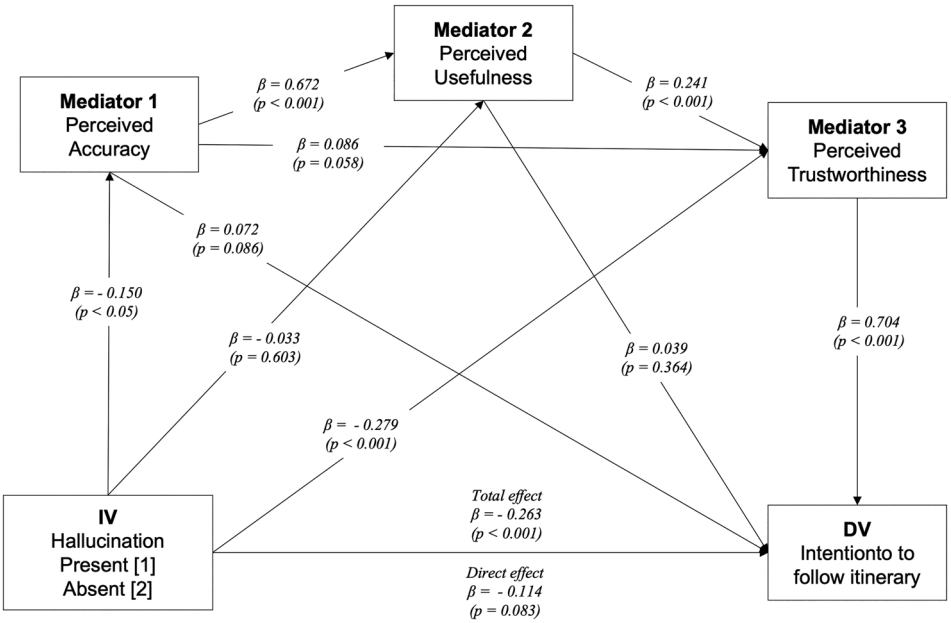

FIGURE 3 | Serial-mediation results of Study 1. *Indirect effect. Hallucination → Perceived Accuracy → Intention to Follow Itinerary, 95% CI [−0.036, −0.000]. Hallucination → Perceived Usefulness → Intention to Follow Itinerary, 95% CI [−0.015, 0.003]. Hallucination → Perceived Trustworthiness → Intention to Follow Itinerary, 95% CI [−0.303, −0.096]. Hallucination → Perceived Accuracy → Perceived Usefulness → Perceived, Trustworthiness → Intention to Follow Itinerary, 95% CI [−0.035, −0.002].

Participants were students from a Spanish university, recruited in person ($n = 1004$; $M_{age} = 25.5$; 51.2% male; 93.2% had prior knowledge of ChatGPT; 72.4% reported using it regularly or several times per week; 53.6% used GenAI to plan trips).

University students are considered a socio-demographically homogeneous and valid group, as they are amongst the primary users of emerging technologies (Anaya-Sánchez et al. 2024; Flavián et al. 2021) and are of particular interest in the study of GenAI (Christensen et al. 2024; Kim et al. 2025). The sample is located in a well-established tourism destination known for its high visitor volume (Paül-i-Agustí 2025). This destination is at the forefront of adopting GenAI within the tourism sector, playing a strategic role in national tourism. Additionally, travellers here are particularly open to utilising tools that enhance the personalisation and optimisation of itineraries and services (Amadeus and Opinium 2025; Phocuswright 2023). This context builds on previous studies that examined the effects of AI in generic, non-tourism or multi-country samples by concentrating on a mature, high-stakes tourism ecosystem where the integration of GenAI and travellers' openness to new technologies are above average. As a result, this setting increases the likelihood of observing significant outcomes from AI-generated errors, offering a rigorous test of the mechanisms linking these errors to traveller behaviour.

Inclusion criteria required prior use of ChatGPT in the last month and at least one leisure trip in the previous year.

Participants were randomly allocated to one of two conditions, hallucination present ($n = 494$) or hallucination absent ($n = 510$), and evaluated ChatGPT outputs during a simulated travel-planning task. Participants were asked to imagine planning a three-day trip to the city of Barcelona, searching for information about places to visit, and using ChatGPT to obtain recommendations (similar to Kim et al. 2025). The stimuli consisted of real ChatGPT screenshots adapted for experimental control. ChatGPT provided three outputs regarding what to visit, where to eat and the best route to reach the 'Sagrada Família' monument in Barcelona. Participants in the incorrect-information condition received additional content: 'The Palau de la Música is not close to Ciutadella, the restaurant "Gourmet" does not exist, and the suggested route is inadvisable due to distance and terrain'. Participants in the no-error condition were shown the same ChatGPT responses but without the added incorrect information (see Figure 2).

After reviewing the stimuli, participants completed a survey. All variables were measured using 7-point Likert scales, adapted from prior works (Kim, Kim, Kim, and Kim 2023; Kim et al. 2025). Finally, participants reported their prior experience with GenAI and provided demographic information.

3.2.2 | Results and Discussion of Study 1

The perceived realism was rated as moderate but significantly above the neutral point ($M = 4.86$, $SD = 1.57$, compared to '4'; $t(1003) = 17.382$, $p < 0.001$). As a manipulation check, perceived general accuracy differed between conditions. Participants in the incorrect-information condition rated ChatGPT as less accurate than those in the no-error condition ($M_{\text{incorrect information present}} = 3.933$, $SD = 1.417$; $M_{\text{incorrect information absent}} = 4.528$, $SD = 1.338$; $t(972) = -6.715$, $p < 0.001$, mean difference $= -0.595$, 95% CI [−0.769, −0.421], Cohen's $d = -0.433$). This confirms that the manipulation successfully reduced perceived accuracy as intended.

We estimated a serial-mediation model (PROCESS Model 6) (Hayes 2017) ((X) Hallucination → (M1) Accuracy → (M2) Usefulness → (M3) Trustworthiness → (Y) Intention to Follow Itinerary) with 5000 bootstrap resamples. Direct paths showed that the presence of hallucinations significantly reduced perceived accuracy (M1; $\beta = -0.150$, $t = 2.071$, $p = 0.038$), accuracy increased usefulness (M2; $\beta = 0.672$, $t = 25.021$, $p < 0.001$), usefulness increased trustworthiness (M3; $\beta = 0.241$, $t = 5.228$, $p < 0.001$) and trustworthiness strongly increased intention to follow the itinerary (Y; $\beta = 0.704$, $t = 24.010$, $p < 0.001$). The M1 (Accuracy) → M3 (Trustworthiness) link was non-significant ($\beta = 0.086$, $p = 0.058$), M1 (Accuracy) → Y (Intention to Follow Itinerary) was non-significant ($\beta = 0.072$, $p = 0.086$) and M2 (Usefulness) → Y (Intention to Follow Itinerary) was non-significant ($\beta = 0.039$, $p = 0.364$). The direct effect of hallucinations on intention was not significant (X (Hallucination) → Y (Intention to Follow Itinerary): $\beta = 0.114$, $t = 1.732$, $p = 0.083$). Notably, X (Hallucination) → M3 (Trustworthiness) was negative ($\beta = -0.279$, $t = 3.756$, $p < 0.001$). Bootstrapped specific indirect effects indicated a significant, negative and significant serial indirect effect (X (Hallucination) → M1 (Accuracy) → M2 (Usefulness) → M3 (Trustworthiness) → Y (Intention to Follow Itinerary)) (95% CI [−0.035, −0.002]), and a small, negative accuracy-only path (X (Hallucination) → M1 (Accuracy) → Y (Intention to Follow Itinerary)) (95% CI [−0.036, −0.000]). The usefulness-only path was not significant (95% CI [−0.015, 0.003]). By contrast, the trust-only path X (Hallucination) → M3 (Trustworthiness) → Y (Intention to Follow Itinerary) was also negative and significant (95% CI [−0.303, −0.096]).

This pattern suggests a suppression: whilst hallucinations lower intention indirectly via reduced accuracy → usefulness → trust (supporting the theorised route), they are also associated with a parallel negative path via trust alone, which helps explain the non-significant total direct effect on intention. Hypotheses H2 (hallucinations ↓ accuracy), H3 (accuracy → usefulness), H4 (usefulness → trust) and H5 (trust → intention) were supported by the significant direct paths reported. Meanwhile, H1 (hallucination ↓ intention) was not supported as a direct effect, but it was supported conditionally through the significant negative serial indirect effect (see Figure 3).

3.3 | Study 2: Testing the Moderating Role of Valence and Salience

Study 2 examined the moderation effects between error valence and error salience, testing whether the observation of negative consequences (e.g., arriving at a closed or non-existent restaurant) amplified the detrimental impact of hallucinations compared to positive outcomes (e.g., discovering an enjoyable attraction despite minor inaccuracies).

3.3.1 | Procedure, Measurement and Participants

Study 2 employed a 2 (hallucination: present vs. absent) × 2 (observed outcome valence: positive vs. negative) between-subjects design ($n = 241$). We define error salience as the conjunction of hallucination present and negative outcome. It is a derived condition of the factorial design rather than an additional factor. Participants first evaluated real ChatGPT screenshots simulating a three-day trip to Barcelona (similar to Study 1). The treatment included (hallucinations) factual inaccuracies (e.g., a non-existent restaurant or an inadvisable route), whereas the control presented identical outputs without errors (see Figure 2). Afterwards, participants read a short message-style vignette describing the outcome of acting on the recommendation (positive vs. negative). Random assignment was applied in both stages. Manipulation checks confirmed (a) detection of erroneous content and (b) correct identification of outcome valence (see Figure 4).

To determine the required sample size for testing the double moderation, we conducted an a priori power analysis in G*Power (Kang 2021). We specified an F test for linear multiple regression (fixed model, R^2 increase), assuming a medium effect size ($f^2 = 0.15$), an alpha level of 0.05, a desired power of 0.95, with two tested predictors and a total of five predictors in the model. The analysis indicated that a minimum of 107 participants would be required to detect the hypothesised effects. Our actual sample exceeded this threshold, thereby ensuring adequate power for detecting the double moderation effects.

As in Study 1, participants rated the same constructs (accuracy, realism and intention to follow the itinerary) on 7-point scales adapted from prior work (Kim, Kim, Kim, and Kim 2023; Kim et al. 2025). Manipulation checks verified (a) detection of erroneous content (present vs. absent) and (b) correct identification of outcome valence (positive vs. negative). Error salience denotes cases in which participants were exposed to a hallucination and observed a negative outcome; it is a descriptive label derived from the 2 × 2 design rather than an additional factor.

The sampling strategy mirrored Study 1 and involved a similar university-student cohort recruited in person at a Spanish university ($N = 241$; $M_{age} = 20.45$; 54% female). The same inclusion criteria were applied: prior use of ChatGPT in the last month and at least one leisure trip in the previous year. Participants were randomly allocated to one of two conditions, hallucination present ($n = 170$) and hallucination absent ($n = 71$); hallucination present/negative ($n = 123$) versus hallucination present/positive ($n = 47$). Most participants reported prior knowledge of ChatGPT (95%); 75% used it regularly or several times per week; 61% were familiar or very familiar with the tool; and 71% indicated they knew how to interact with ChatGPT.

3.3.2 | Results and Discussion of Study 2

Perceived realism was rated as moderate but significantly above the neutral point ($M = 4.71$, SD = 1.39, compared to '4'; $t(229) = 18.72$, $p < 0.001$). A one-way ANOVA was conducted on the key metrics for the main analysis.

The experimental factor had a significant effect on the perceived accuracy of the information provided. Specifically, results indicated that perceived accuracy was higher when incorrect information was absent ($M_{absent} = 4.85$, SD = 1.42; compared to $M_{present} = 3.28$, SD = 1.78; $F(1, 239) = 43.61$, $p < 0.001$, $\eta^2 = 0.154$). Thus, the perceived accuracy of the information provided by ChatGPT was greater in the absence of hallucinated content.

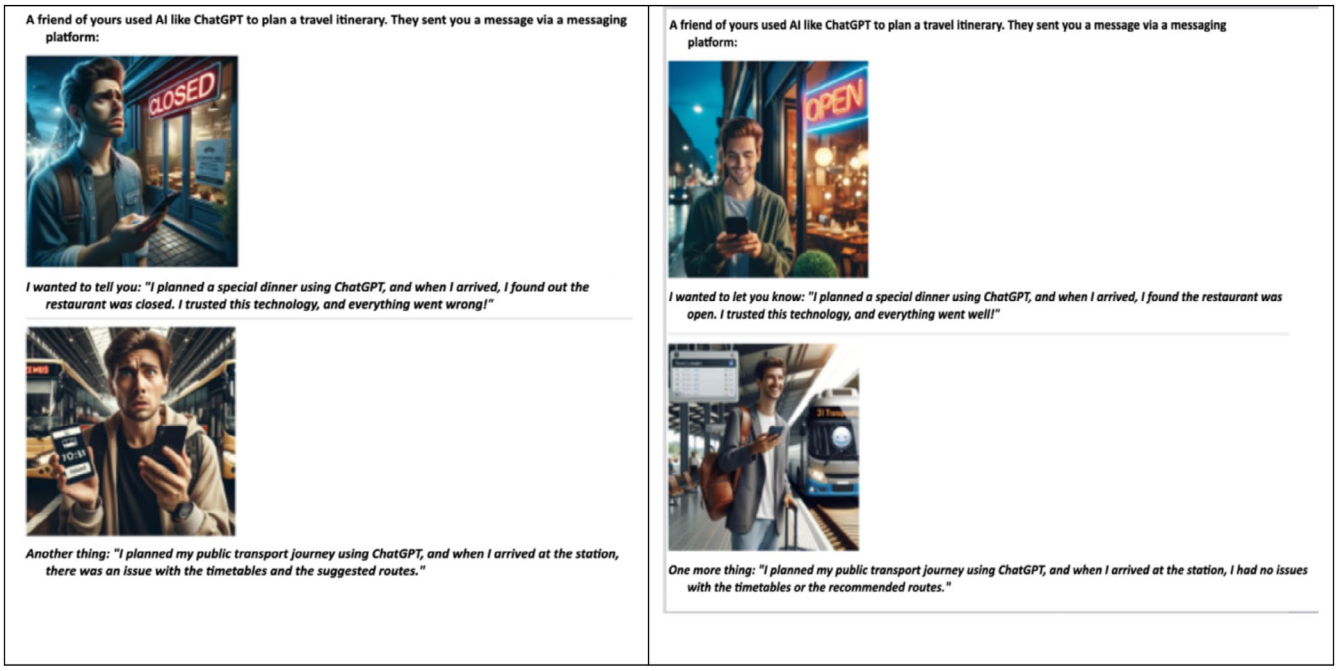

FIGURE 4 | Stimuli for Study 2. Participants viewed two message-style vignettes about a friend who planned a trip using ChatGPT. *Left panel*: (hallucination present) negative outcome caused by AI factual inaccuracies, on arrival, the restaurant was closed, and the public transport itinerary had timetable/route issues. *Right panel*: (hallucination absent) positive outcome with an open restaurant and a smooth public transport journey. Images are illustrative; the text and layout are identical across panels except for outcome valence.

To test H6, we asked whether what happens after following ChatGPT's advice, a positive versus a negative outcome, changes how much an error affects people's willingness to follow the itinerary. As a first step, we simply compared participants' intentions across the positive-outcome and negative-outcome conditions. A one-way ANOVA showed a robust effect, $F(1, 239) = 70.81$, $p < 0.001$, $\eta^2 = 0.23$. Participants reported higher intention under a positive outcome (i.e., when no negative consequence followed the hallucinations; $M = 4.94$, SD = 1.30) than under a negative outcome ($M = 3.33$, SD = 1.50). This pattern is consistent with H6, indicating that adverse consequences magnify the detrimental effect of errors on compliance with the itinerary.

To test H7, this study examines the moderating role of error salience in the incorrect information generated by ChatGPT on participants' intention to follow its recommendations. To test this hypothesis, a 2 (incorrect information: present vs. absent) × 2 (error outcome: negative vs. positive) two-way factorial ANOVA was conducted within a repeated measures design. Results revealed a significant within-subject effect related to the observation of a negative outcome ($F(1, 238) = 6.55$, $p = 0.011$, $\eta^2 = 0.01$). Post hoc comparisons indicated that participants exposed to a negative outcome expressed significantly lower intention to follow the recommendations (mean difference = −0.29, $p = 0.011$). Regarding between-subjects effects, a significant main effect was identified ($F(2, 238) = 46.10$, $p < 0.001$, $\eta^2 = 0.22$). Post hoc analyses revealed significant differences between the groups. Participants repeatedly exposed to hallucinations and who observed negative consequences showed a lower intention to follow ChatGPT's advice (mean difference = 1.93, $p < 0.001$). Additionally, a significant interaction was observed between the type of outcome (positive vs. negative) and error salience ($F(2, 238) = 1.71$, $p = 0.183$, $\eta^2 = 0.00$). However, post hoc comparisons showed that participants exposed to hallucinations and a negative outcome (salient incorrect information) reported significantly lower intention to follow the recommendations compared to other groups (mean difference up to −2.13, $p < 0.001$). This finding supports H7. These results confirm that the visibility of incorrect information generated by ChatGPT has a particularly strong negative effect when users directly observe the adverse consequences of such errors. This study provides new evidence that the negative impact of ChatGPT's error visibility is significantly amplified when users witness tangible adverse outcomes resulting from those mistakes.

4 | General Discussion and Implications

First, Study 1 shows that hallucinations work through a simple sequence: lower accuracy reduces usefulness, which then undermines trust, and this pathway, in turn, shapes intention to follow the itinerary; thus, we clarify the mechanism that earlier work left at the level of acceptance and trust outcomes (Christensen et al. 2024; Kim, Kim, Park, et al. 2023; Kim et al. 2025).

H1 proposed that hallucinations directly reduce the intention to follow an itinerary. However, no significant direct effect was found. This finding is important because it indicates that hallucinations are not harmless; rather, their negative impact operates in a more complex manner than initially expected, flowing through a series of cognitive and attitudinal evaluations. These results contribute to earlier studies by Kim et al. (2025) and Yaprak (2024), which did identify direct effects on intentions. Our model suggests that examining only the direct relationship may overlook the underlying mechanisms responsible for it.

Accordingly, our evidence clarifies mixed findings in the literature by showing that the impact of hallucinations on intention is predominantly indirect and mechanism-bound rather than uniformly direct.

H2 posits that the presence of hallucinations negatively impacts the perceived accuracy of recommendations generated by ChatGPT. This hypothesis was empirically supported, confirming that hallucinations significantly diminish the perceived accuracy of these recommendations. Hallucinations can be seen as a critical point of failure in the relationship between users and the information presented, as they represent factual inaccuracies and indicate that the information quality, serving as the initial filter for users, has failed. This finding aligns with established technology acceptance models, such as the TAM and the TPB (Ajzen 1985; Davis 1989), as well as existing literature on technological trust (e.g., McKnight et al. 2011), which emphasises that the credibility of information is a fundamental prerequisite. In relation to H3, it was confirmed that a higher perceived accuracy enhances the perceived usefulness of ChatGPT-generated recommendations. This connection is logical, as incorrect information (characterised by low accuracy) cannot be helpful in a travel-planning context, where the consequences of errors can be significant. Therefore, our findings provide empirical support for the relationship posited by Kim, Kim, Kim, and Kim (2023) and Park et al. (2007), who assert that factual correctness is essential for users to deem a recommendation valuable.

H4 received empirical support, demonstrating that perceived usefulness positively influences users' trust in ChatGPT. When users see a GenAI system as helpful in achieving their travel goals, trust begins to develop. This finding aligns with the research of Hrankai and Mak (2025) and Demir and Demir (2023), who identify system performance and value creation as key factors that foster trust in GenAI.

Regarding H5, trustworthiness emerged as a strong and positive predictor of users' intention to follow the proposed itinerary. This relationship marks the transition where trust, an attitudinal factor, translates into behaviours, in this case, the intention to follow the itinerary. The literature on trust in GenAI within the tourism sector supports this idea, emphasising trust as the central attitudinal variable that drives acceptance (Kim et al. 2025; Loureiro et al. 2024).

These findings reinforce the notion that the perception of hallucinations does not operate in isolation but is integrated within a process involving the evaluation of the experience itself. As suggested in previous research (Kim et al. 2025; Yaprak 2024), our findings indicate that itinerary errors have a significant impact on behavioural intentions. These findings support previous models of technological trust (Filieri and McLeay 2014; McKnight et al. 2011) by demonstrating that acceptance of GenAI-generated recommendations depends on a complex cognitive structure in which credibility is progressively built or undermined through the interaction of accuracy, usefulness and trust. Particularly relevant in this regard, this highlights the need to differentiate these constructs when analysing the impact of GenAI errors. Our findings contribute to addressing the underexplored 'dark side' of GenAI in consumer behaviour, a research stream increasingly recognised as vital for understanding the broader societal impact of disruptive technologies (Beheshti et al. 2023).

In Study 2, the boundary conditions regarding the impact of hallucinations on trust and intention, finding that the damage inflicted is not consistent across situations. The study aimed to determine when this negative impact intensifies by examining users' concrete experiences after they acted on AI-generated recommendations. Supporting H6, the results revealed that the valence of the observed outcome significantly influences the acceptance of the recommendation. Participants who experienced a positive outcome (e.g., the recommended restaurant was open and their experience was enjoyable) reported a much higher intention to follow the recommendation compared to those who encountered a negative outcome (e.g., the restaurant was closed or did not exist). This finding is crucial as it demonstrates that a tangible negative outcome greatly amplifies the negative effects of a hallucination. Conversely, a positive outcome, even if it results from serendipity, meaning the outcome is unexpectedly valuable, can mitigate these effects. This aligns with post-use trust models (McKnight et al. 2002) and experiential learning theories, which propose that concrete experiences shape future evaluations.

Furthermore, H7 focused on the prominence of errors. The results supported this hypothesis, showing that when errors were prominent, intentions to follow the recommendations dropped significantly. This finding indicates the tipping point at which the impact of the hallucination becomes decisively negative for the user. Theoretically, these findings connect directly with the accessibility-diagnosticity framework (Feldman and Lynch Jr. 1988). A technical error (such as a hallucination) that does not lead to negative consequences might be overlooked by the user due to its low diagnosticity. However, when the hallucination results in a tangible negative experience (like a non-existent restaurant or an incorrect schedule that disrupts plans), that failure becomes highly memorable and indicative of the system's overall reliability.

Finally, these results provide empirical support for the concept of algorithm aversion (Dietvorst et al. 2015). The literature suggests that consumers tend to evaluate errors in automated systems more harshly than human errors, particularly when these errors lead to negative outcomes. Study 2 illustrates this phenomenon: a visible AI error is not simply regarded as a failure but as a violation of the high expectations for accuracy placed on GenAI. This intensifies the consumer's rejection response.

4.1 | Theoretical Implications

The findings from Studies 1 and 2 suggest that hallucinations primarily undermine acceptance through a series of cognitive evaluations and attitude appraisals, and that negative experiences following their use exacerbate this process (Christensen et al. 2024; Kim et al. 2025). According to the accessibility-diagnosticity framework (Feldman and Lynch Jr. 1988), noticeable errors are interpreted as indicators of the overall performance of the system. Compared to other sources, such as official websites or user reviews, GenAI tends to create higher expectations for accuracy. When these expectations are not

met, the resulting disappointment amplifies rejection responses and strengthens the tendency for algorithm aversion (Dietvorst et al. 2015).

This research contributes to the literature on trust in GenAI and consumer behaviour in digital environments by integrating cognitive and attitudinal effects of hallucinations with experiential and perceptual factors in the context of travel planning. The study proposes and validates a serial-mediation model that demonstrates how perceived accuracy, usefulness and trust influence users' responses to hallucinations generated by GenAI technology (ChatGPT).

The positioning of trust within the TAM and the TPB has been refined by recent evidence. This research indicates that perceived accuracy is a prerequisite for perceived usefulness, demonstrating that trust serves as a crucial attitudinal link between usefulness and intention in the context of GenAI. Particularly in time-sensitive and consequence-rich tasks, such as itinerary planning, accuracy plays a vital role in shaping judgements of usefulness.

In relation to TPB, trust can be understood as a specific attitudinal belief regarding the reliability of the GenAI source, which influences performance beliefs and ultimately, intention. Additionally, the outcomes observed can further modify behavioural beliefs and enhance perceived capability to depend on the system.

Consequently, the findings suggest two important pathways: one consistent with TAM, where accuracy enhances usefulness, usefulness builds trust and trust drives intention, and another aligned with TPB, where belief revision based on concrete experiences shifts intentions in uncertain situations (Ajzen 1985; Davis 1989; McKnight et al. 2002). In summary, we define accuracy as an upstream belief within the TAM and formalise trust as the key factor that translates perceived usefulness into behavioural intention. Additionally, the TPB is enhanced by incorporating experience-driven belief revision as a boundary mechanism in situations of uncertainty.

This study extends the work of Kim et al. (2025), who focused on the main effect of errors, and demonstrates that error salience and, especially, the experienced outcome after following the advice markedly magnify rejection. Furthermore, we complement Christensen et al. (2024) by shifting from familiarity-based predictors to a validated serial-mediation model, and we add to work centred on hallucination typologies and measurement (Ji et al. 2023; Sun et al. 2024) with empirical evidence tracing the perception-to-behaviour route and showing that salient, negative outcomes are the tipping point. It also identifies the moderating role of observed experience and error salience, expanding understanding of the mechanisms that shape trust in these technologies.

4.2 | Practical Implications

From a managerial perspective, our findings transform GenAI hallucinations from a technical error into a strategic risk to trust for tourism companies. Travellers often tend to accept seemingly neutral GenAI advice, and the rapidly changing information about destinations creates a fertile ground for misinformation. In this context, digital trust becomes a scarce competitive asset. From a management perspective, our findings indicate that merely 'eliminating hallucinations' is not enough. Firms must raise perceived accuracy and trust simultaneously. Product roadmaps should prioritise features that (a) strengthen factual grounding (to improve accuracy), (b) surface decision-support value (to enhance usefulness) and (c) make reliability auditable (to build trust). Doing so targets precisely the links identified in our serial pathway. Accordingly, firms should institutionalise hallucination management as a core capability: map all GenAI touchpoints (pre-trip search, itinerary planning, booking, in-trip assistance), assess their exposure to hallucinations and prioritise mitigation where the stakes are highest. For high-risk communications, prices and fees, booking confirmations, safety and entry requirements, adopt a 'human-over-the-loop' workflow so that critical outputs are verified by staff before reaching customers. Organisations that lead with reliable AI will compound loyalty, whereas laggards risk accelerated erosion of trust in an increasingly uncertain information ecosystem.

Second, those levers should be translated into system design choices. To improve perceived accuracy at scale, platforms should favour retrieval-augmented generation (RAG) that grounds answers in verified, real-time sources (e.g., official attractions data, transport timetables) rather than model-only generation. This is directly aimed at the validated route: accuracy → usefulness → trustworthiness. To curb over-reliance and the 'black-box' perception, implement automated validation and anomaly alerts, display clear GenAI labels, and add interface nudges that prompt users to verify time-sensitive facts independently (Christensen et al. 2024; Sun et al. 2024). Visible disclaimers, source citations, optional 'show evidence' views and conservative abstention policies (e.g., 'I do not know, click to check official site') can further reduce unwarranted confidence. Establish feedback channels for error reporting and close the loop by triaging, fixing and communicating corrections. Because perceived information quality strongly shapes trust and brand relationships online, continuous improvements to accuracy and clarity should be treated as both a technical goal and an engagement strategy (see Christensen et al. 2024; Nguyen et al. 2022; Sun et al. 2024).

Third, manage the experiential amplifiers our Study 2 uncovered. Outcome valence had a significant effect on intention, and intentions fell further when error salience was high. Practically, triage for salient-risk scenarios, venues with volatile hours, new restaurants, strike-prone transit and add extra safeguards: real-time open/closed checks before suggesting a venue, geofenced 'last-mile' validations in route, conservative route plans with verified alternates and graceful-degradation policies (e.g., 'cannot verify, here are two confirmed options'). Pair this with service-recovery playbooks (instant re-routing, vouchers, apologies) when negative outcomes occur because these episodes disproportionately damage trust, and rapid recovery limits knock-on effects on future intention. Operationally, we recommend formal governance with measurable trust KPIs: hallucination rate (and severity-weighted incident rate), verified-source coverage, negative-outcome rate (missed/closed venues, failed bookings) and user-reported

trust. Invest in organisation-wide GenAI literacy so front-line and product teams understand model limits and practice verification habits, not just tool usage.

Finally, we note a longer-term constraint: durable mitigation likely requires socio-technical change in how models are evaluated and rewarded. As Kalai et al. (2025) argue, if leaderboards penalise uncertainty and reward confident guessing, models remain incentivised to hallucinate. Whilst that ecosystem shift unfolds, tourism firms can still act now, ground outputs, verify high-stakes claims, make uncertainty visible and cultivate user feedback loops to deliver more explainable, dependable and user-centred GenAI services.

4.3 | Limitations and Opportunities for Future Research

Finally, this study has certain limitations. Both samples consisted mainly of university students, limiting the findings' generalisability. Moreover, using simulated scenarios restricts the ability to capture the complexity of real-life interactions in dynamic environments fully. Future research should replicate these studies with more heterogeneous samples, assess interactions in natural contexts, incorporate cultural variables and examine the effect of hallucination mitigation strategies such as disclaimers or the presentation of alternative recommendations. It would also be relevant to explore how trust in GenAI evolves over repeated exposures or over time.

5 | Conclusion

This research reveals that GenAI hallucinations in tourism are serious risks that undermine consumer trust and intentions through a multi-stage cognitive process, especially when negative outcomes highlight errors. Our two-study design confirms a serial-mediation model and provides practical guidance for tourism companies. It emphasises the need to manage GenAI hallucinations as a key organisational capability, combining technical solutions with governance and service recovery. Overall, the findings contribute to discussions about how disruptive GenAI technologies influence consumer behaviour, trust and decision-making in digital markets, exposing both vulnerabilities and opportunities in tourism.

Acknowledgements

This work was supported by the Funding for open access charge was provided by Universidad de Malaga/CBUA. Additionally, funding was provided by the Institutional project B3-2023/21-(MP-ECSTIA) and by the project 'Chronic Emergencies and Ecosocial Transformation in Turistified Litoral Spaces' (PID2022-137648OB-C22) financed by the Spanish Ministry of Science and Innovation (National Plan for R + D + i), the Spanish State Research Agency and the European Regional Development Fund (ERDF). University of Malaga's II Plan for Research, Knowledge Transfer and Dissemination (call 2023), with funding from the Autonomous Community of Andalusia (Spain) and FEDER Andalusia Programme 2021-2027 (Grant: PAIDIResearchGroupSEJ-567).

Funding

This work was supported by the Funding for open access charge was provided by Universidad de Malaga/CBUA, Institutional project – University of Málaga (Grant No. B3-2023/21-(MP-ECSTIA)) and Spanish Ministry of Science and Innovation (National Plan for R + D + i), the Spanish State Research Agency and the European Regional Development Fund (ERDF) (Grant No. PID2022-137648OB-C22). University of Malaga's II Plan for Research, Knowledge Transfer and Dissemination (call 2023), with funding from the Autonomous Community of Andalusia (Spain) and FEDER Andalusia Programme 2021-2027 (Grant: PAIDI ResearchGroupSEJ-567).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Ajzen, I. 1985. "From Intentions to Actions: A Theory of Planned Behavior." In *Action Control: From Cognition to Behavior*, edited by J. Kuhl and J. Beckmann, 11–39. Springer. https://doi.org/10.1007/978-3-642-69746-3_2.

Ali, F., B. Yasar, L. Ali, and S. Dogan. 2023. "Antecedents and Consequences of Travelers' Trust Towards Personalized Travel Recommendations Offered by ChatGPT." *International Journal of Hospitality Management* 114: 103588.

Amadeus, and Opinium. 2025. "Connected Journeys: How Will Technology Transform Travel in the Next Decade." https://amadeus.com/content/dam/amadeuswebevo/brand/amadeus/photos/sections/transforming-travel/amadeus-connected-journeys-1.pdf.

Anaya-Sánchez, R., F. Rejón-Guardia, and S. Molinillo. 2024. "Impact of Virtual Reality Experiences on Destination Image and Visit Intentions: The Moderating Effects of Immersion, Destination Familiarity and Sickness." *International Journal of Contemporary Hospitality Management* 36, no. 11: 3607–3627. https://doi.org/10.1108/IJCHM-09-2023-1488.

Beheshti, M., A. ZareRavasan, H. A. Mahdiraji, V. Jafari-Sadeghi, and G. Sakka. 2023. "An Overview of the Consumer-Centric Disruptive Technology Research: Insights From Topic Modelling and Literature Review." *Journal of Consumer Behaviour* 23, no. 2: 372–388. https://doi.org/10.1002/cb.2204.

Carvalho, F., and S. Ivanov. 2023. "ChatGPT for Tourism: Applications, Benefits and Risks." *Tourism Review* 79, no. 2: 290–303. https://doi.org/10.1108/TR-02-2023-0088.

Chang, W., and J. Park. 2024. "A Comparative Study on the Effect of ChatGPT Recommendation and AI Recommender Systems on the Formation of a Consideration Set." *Journal of Retailing and Consumer Services* 78: 103743. https://doi.org/10.1016/j.jretconser.2024.103743.

Christensen, J., J. M. Hansen, and P. Wilson. 2024. "Understanding the Role and Impact of Generative Artificial Intelligence (AI) Hallucination Within Consumers' Tourism Decision-Making Processes." *Current Issues in Tourism* 28, no. 4: 545–560. https://doi.org/10.1080/13683500.2023.2300032.

Davis, D. 1989. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology." *MIS Quarterly* 13, no. 3: 319–340. https://doi.org/10.2307/249008.

Demir, M., and S. S. Demir. 2023. "Is ChatGPT the Right Technology for Service Individualization and Value?" *Journal of Service Theory and*

Practice 33, no. 9: 1184–1206. https://doi.org/10.1080/10548408.2023.2255884.

Dietvorst, B. J., J. P. Simmons, and C. Massey. 2015. "Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err." *Journal of Experimental Psychology: General* 144, no. 1: 114–126. https://doi.org/10.1037/xge0000033.

Duong, C. D., T. H. Nguyen, T. V. N. Ngo, T. T. P. Pham, A. T. Vu, and N. S. Dang. 2025. "Using Generative Artificial Intelligence (ChatGPT) for Travel Purposes: Parasocial Interaction and Tourists' Continuance Intention." *Tourism Review* 80, no. 4: 813–827. https://doi.org/10.1108/TR-01-2024-0027.

Feldman, J. M., and J. G. Lynch Jr. 1988. "Self-Generated Validity and Other Effects of Measurement on Belief, Attitude, Intention, and Behavior." *Journal of Applied Psychology* 73, no. 3: 421–435. https://doi.org/10.1037/0021-9010.73.3.421.

Filieri, R., and F. McLeay. 2014. "E-WOM and Accommodation: An Analysis of the Factors That Influence Travellers' Adoption of Information From Online Reviews." *Journal of Travel Research* 53, no. 1: 44–57. https://doi.org/10.1177/0047287513481274.

Flavián, C., S. Ibáñez-Sánchez, and C. Orús. 2021. "Impacts of Technological Embodiment Through Virtual Reality on Potential Guests' Emotions and Engagement." *Journal of Hospitality Marketing & Management* 30, no. 1: 1–20. https://doi.org/10.1080/19368623.2020.1770146.

Grewal, D., A. Guha, C. B. Satornino, and E. B. Schweiger. 2021. "Artificial Intelligence: The Light and the Darkness." *Journal of Business Research* 136: 229–236. https://doi.org/10.1016/j.jbusres.2021.07.043.

Gursoy, D., Y. Li, and H. Song. 2023. "ChatGPT and the Hospitality and Tourism Industry: An Overview of Current Trends and Future Research Directions." *Journal of Hospitality Marketing & Management* 32, no. 5: 579–592. https://doi.org/10.1080/19368623.2023.2211993.

Guttentag, D. A., S. W. Litvin, and R. Teixeira. 2024. "Human vs. AI: Can ChatGPT Improve Tourism Product Descriptions?" *Current Issues in Tourism* 28: 1–19. https://doi.org/10.1080/13683500.2024.2402563.

Hayes, A. F. 2017. *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. Guilford Publications.

Hrankai, R., and B. Mak. 2025. "Bridging the Affordance-Actualization Gap in User Preferences for AI-Assisted Trip Planning." *Journal of Travel Research*: 1–17. https://doi.org/10.1177/00472875251322518.

Ji, Z., N. Lee, R. Frieske, et al. 2023. "Survey of Hallucination in Natural Language Generation." *ACM Computing Surveys* 55, no. 12: 1–38. https://doi.org/10.1145/3571730.

Jia, Y., A. Garg, and J. Dias Soeiro. 2024. "A Comprehensive Framework for Understanding Consumers' Intentions of Artificial Intelligence Devices in the Hospitality Industry: A Meta-Analysis." *Current Issues in Tourism*: 1–19. https://doi.org/10.1080/13683500.2024.2429733.

Kalai, A. T., O. Nachum, S. S. Vempala, and E. Zhang. 2025. "Why Language Models Hallucinate". Preprint, arXiv:2509.04664. https://doi.org/10.48550/arXiv.2509.04664.

Kalai, A. T., and S. S. Vempala. 2024. "Why Language Models Hallucinate." Preprint, arXiv:2509.04664. https://doi.org/10.48550/arXiv.2509.04664.

Kang, H. 2021. "Sample Size Determination and Power Analysis Using the G*Power Software." *Journal of Educational Evaluation for Health Professions* 18: 1–12. https://doi.org/10.3352/jeehp.2021.18.17.

Kim, J. H., J. Kim, C. Kim, and S. Kim. 2023. "Do You Trust ChatGPTs? Effects of the Ethical and Quality Issues of Generative AI on Travel Decisions." *Journal of Travel & Tourism Marketing* 40, no. 9: 779–801. https://doi.org/10.1080/10548408.2023.2293006.

Kim, J. H., J. Kim, S. Kim, and T. B. Hailu. 2023. "Effects of AI ChatGPT on Travelers' Travel Decision-Making." *Tourism Review* 79, no. 5: 1038–1057. https://doi.org/10.1108/TR-07-2023-0489.

Kim, J. H., J. Kim, J. Park, C. Kim, J. Jhang, and B. King. 2023. "Decisions With ChatGPT: Reexamining Choice Overload in ChatGPT Recommendations." *Journal of Retailing and Consumer Services* 75: 103494. https://doi.org/10.1016/j.jretconser.2023.103494.

Kim, J. H., J. Kim, J. Park, C. Kim, J. Jhang, and B. King. 2025. "When ChatGPT Gives Incorrect Answers: The Impact of Inaccurate Information by Generative AI on Tourism Decision-Making." *Journal of Travel Research* 64, no. 1: 51–73. https://doi.org/10.1177/00472875231212996.

Kumar, M., U. A. Mani, P. Tripathi, et al. 2023. "Artificial Hallucinations by Google Bard: Think Before You Leap." *Cureus* 15, no. 8: e43313. https://doi.org/10.7759/cureus.43313.

Liu, Y., and Y. Du. 2025. "The Effect of Generative AI Ethics on Users' Continuous Usage Intentions: A PLS-SEM and fsQCA Approach." *International Journal of Human-Computer Interaction* 41: 1–12. https://doi.org/10.1080/10447318.2025.2465861.

Loureiro, S. M. C., J. Guerreiro, and P. Rita. 2024. "Tourists and Artificial Intelligence-LLM Interaction: The Power of Forgiveness." *Current Issues in Tourism* 28, no. 7: 1172–1190. https://doi.org/10.1080/13683500.2024.2353872.

McKnight, D. H., M. Carter, J. B. Thatcher, and P. F. Clay. 2011. "Trust in a Specific Technology: An Investigation of Its Components and Measures." *ACM Transactions on Management Information Systems* 2, no. 2: 1–25. https://doi.org/10.1145/1985347.1985353.

McKnight, D. H., V. Choudhury, and C. Kacmar. 2002. "Developing and Validating Trust Measures for e-Commerce: An Integrative Typology." *Information Systems Research* 13, no. 3: 334–359. https://doi.org/10.1287/isre.13.3.334.81.

Nguyen, T. M., S. Quach, and P. Thaichon. 2022. "The Effect of AI Quality on Customer Experience and Brand Relationship." *Journal of Consumer Behaviour* 21, no. 3: 481–493. https://doi.org/10.1002/cb.1974.

OpenAI. 2024. "Why Language Models Hallucinate." [Blog Post]. https://openai.com/research/why-language-models-hallucinate.

OpenAI. 2025. "Why Language Models Hallucinate." https://openai.com/index/why-language-models-hallucinate/.

Park, D.-H., J. Lee, and I. Han. 2007. "The Effect of On-Line Consumer Reviews on Consumer Purchasing Intention: The Moderating Role of Involvement." *International Journal of Electronic Commerce* 11, no. 4: 125–148. https://doi.org/10.2753/JEC1086-4415110405.

Paül-i-Agustí, D. 2025. "The Concentrated City: Effects of AI-Generated Travel Advice on the Spatial Distribution of Tourists." *Urban Science* 9, no. 7: 268. https://doi.org/10.3390/urbansci9070268.

Phocuswright. 2023. "GenAI and Emerging Tech Use Among Travelers." https://www.phocuswright.com/Travel-Research/Consumer-Trends/GenAI-and-Emerging-Tech-Use-Among-Travelers.

Pretolesi, D., O. Zechner, H. Schrom-Feiertag, and M. Tscheligi. 2024. "Can I Trust You? Exploring the Impact of Misleading AI Suggestions on User's Trust." In *2024 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*, 1230–1235. https://doi.org/10.1109/MetroXRAINE62247.2024.10796402.

Puntoni, S., R. W. Reczek, M. Giesler, and S. Botti. 2021. "Consumers and Artificial Intelligence: An Experiential Perspective." *Journal of Marketing* 85, no. 1: 131–151. https://doi.org/10.1177/0022242920953847.

Rawte, V., S. Chakraborty, A. Pathak, et al. 2023. "The Troubling Emergence of Hallucination in Large Language Models – An Extensive Definition, Quantification, and Prescriptive Remediations." Preprint, arXiv:2310.04988. https://doi.org/10.48550/arXiv.2310.04988.

Shi, S., Y. Gong, and D. Gursoy. 2021. "Antecedents of Trust and Adoption Intention Toward Artificially Intelligent Recommendation Systems in Travel Planning: A Heuristic–Systematic Model." *Journal of Travel Research* 60, no. 8: 1714–1734. https://doi.org/10.1177/0047287520966395.

Shin, S., J. Kim, E. Lee, Y. Yhee, and C. Koo. 2025. "ChatGPT for Trip Planning: The Effect of Narrowing Down Options." *Journal of Travel Research* 64, no. 2: 247–266.

Solomovich, L., and V. Abraham. 2024. "Exploring the Influence of ChatGPT on Tourism Behavior Using the Technology Acceptance Model." *Tourism Review*: 1–18. https://doi.org/10.1108/TR-10-2023-0697.

Sun, Y., D. Sheng, Z. Zhou, and Y. Wu. 2024. "AI Hallucination: Towards a Comprehensive Classification of Distorted Information in Artificial Intelligence-Generated Content." *Humanities and Social Sciences Communications* 11: 1278. https://doi.org/10.1057/s41599-024-03811-x.

Topsakal, Y. 2025. "How Familiarity, Ease of Use, Usefulness, and Trust Influence the Acceptance of Generative Artificial Intelligence (AI)-Assisted Travel Planning." *International Journal of Human-Computer Interaction* 41, no. 15: 9478–9491. https://doi.org/10.1080/10447318.2024.2426044.

van Dis, E. A., J. Bollen, W. Zuidema, R. Van Rooij, and C. L. Bockting. 2023. "ChatGPT: Five Priorities for Research." *Nature* 614, no. 7947: 224–226. https://doi.org/10.1038/d41586-023-00288-7.

Wong, I. A., Q. L. Lian, and D. Sun. 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.

Xu, H., X. Li, J. C. Lovett, and L. T. O. Cheung. 2025. "ChatGPT for Travel-Related Services: A Pleasure–Arousal–Dominance Perspective." *Tourism Review* 80, no. 5: 901–922. https://doi.org/10.1108/TR-07-2024-0570.

Xu, J., X. Tang, H. Lin, and Y. Luo. 2024. "The Impact of Recommender Type on Tourist Price Sensitivity to Travel Plans." *Journal of Travel Research* 64, no. 7: 1680–1700. https://doi.org/10.1177/00472875241261633.

Yaprak, B. 2024. "Generative Artificial Intelligence in Marketing: The Invisible Danger of AI Hallucinations." *Journal of Economy Business and Management* 8, no. 2: 133–158. https://doi.org/10.7596/jebm.1588897.

Appendix A

Comparative Literature Matrix on GAI Hallucinations in Tourism Marketing Research

Authors	Research context	Variables	Hypotheses	Sample type and size	Main findings	Future directions and gaps
Kim, Kim, Kim, and Kim (2023)	Impact of quality and ethical concerns on travel decisions with ChatGPT	IVs: quality salience (positive vs. negative), ethical salience (positive vs. negative). Mediator: perceived trustworthiness. Moderators: ChatGPT errors (present/absent), information type (general vs. specific). DVs: visit intention, satisfaction, acceptance.	Errors moderate the negative impact of low quality and ethical concerns; tolerance differs by information type.	Four experimental studies with US adults (Amazon MTurk). Study 1: 302 participants. Study 2: 312 participants. Study 3: 382 participants. Study 4: 526 participants.	Negative quality and ethical cues reduced visit intention, satisfaction and acceptance. Perceived trustworthiness mediated negative effects. Errors reduced the negative impact of ethical and quality issues (moderation confirmed). General information strengthened the effect of ethical aspects, indicating less moral decoupling than specific information.	Hypothetical scenarios limit external validity. Conducted in 2023 with GPT-3.5; results may differ with later models. Longitudinal and field studies are needed to validate perception changes. Explore additional moderators (time, cultural differences, cognitive style). Future work: other contexts (e.g., financial planning, health).
Kim, Kim, Kim, and Hailu (2023)	Intentions to use ChatGPT in tourism decisions. Effect of experience, errors and information valence	IVs: prior experience, exposure to errors, information valence. Mediators: perceived usefulness, perceived ease of use. DV: intention to use.	Experience increases intention to use; usefulness and ease of use mediate relationships; errors reduce intention; positive information increases intention.	Four empirical studies; total $n = 1689$ (US, Amazon MTurk).	Prior experience increases intention to use; mediation confirmed; errors reduce intention; positive information increases intention.	Need for behavioural measures and field studies; include more destinations and formats; examine AI evolution.
Christensen et al. (2024)	AI hallucinations and their effect on trust and tourism decisions; TAM and TPB frameworks	IVs: prior use, familiarity. TAM/TPB constructs: perceived ease of use (PEOU), perceived usefulness (PU), attitude, perceived behavioural control (PBC). DV: behavioural intention. Moderator: perceived hallucination potential.	Familiarity with AI increases PEOU, PU, PBC and concern about hallucinations; concern moderates the PEOU–PBC–intention relationship.	An online survey of 900 consumers (North America and Asia–Pacific) was conducted.	All hypotheses were supported; familiarity and prior use drive PEOU, PU and PBC; many participants still chose itineraries containing errors; and concern did not correlate with actual detection.	Lack of external variables; gap between awareness and detection capability; explore UTAUT; consider mobile apps and offline trust.

Authors	Research context	Variables	Hypotheses	Sample type and size	Main findings	Future directions and gaps
Loureiro et al. (2024)	Tourists' attachment–aversion towards AI-LLMs; role of forgiveness after hallucinations	IVs: symbolic, hedonic and functional benefits. Mediator: attachment–aversion. DV: motivational strength. Moderator: forgiveness.	Attachment–aversion affects motivational strength; forgiveness moderates the benefits–attachment link.	Qualtrics survey, $n = 451$.	Symbolic benefits are key to attachment–aversion; forgiveness strengthens relationships; empathetic recovery protocols are recommended.	Limited external validity; lack of real interaction; test experimental settings.
Kim et al. (2025)	Impact of incorrect ChatGPT information on acceptance of travel recommendations; accessibility-diagnosticity theory	IV: incorrect information. Mediators: accuracy, trustworthiness. Moderators: prominence, error type, and initial focus. DV: visit intention.	Incorrect information reduces trust and intention; mediation via accuracy and trustworthiness; moderated by prominence, error type and initial focus.	Six between-subjects experiments (US and UK), participants via MTurk, CloudResearch and Prolific. Samples ≥ 210 per study (determined with G*Power). Approx. total > 1500.	Incorrect information reduces visit intention and trust; greater impact when the error is prominent or in the same domain; initial selection reduces subsequent impact.	Varying error awareness; need for real behaviours; broaden to other contexts; update to recent models; explore mitigation and ethics.
This research	Travel planning with ChatGPT using real outputs (screenshots) and consequence vignettes to capture post-use experiences; integrates accessibility-diagnosticity and technology-trust frameworks into a process-plus-conditions model.	IV: hallucination presence (present/absent). Serial mediators: perceived accuracy → perceived usefulness → trustworthiness. DV: intention to follow the itinerary. Moderators (Study 2): observed outcome valence (positive/negative), prominence (hallucination + negative outcome).	H1 ↓ intention due to hallucination; H2 hallucination ↓ accuracy; H3 accuracy → usefulness; H4 usefulness → trustworthiness; H5 trustworthiness → intention; H6 observed valence moderates (positive attenuates/negative amplifies); H7 error prominence amplifies the effect.	Two controlled experiments (Spain). Study 1: $n = 1004$ (serial mediation). Study 2: 2 × 2 (hallucination × valence), $n = 241$ (moderations; prominence = hallucination + negative outcome). University sample with recent ChatGPT use and at least one leisure trip in the last year.	Novelty 1 (mechanism): evidence of negative serial mediation the direct effect on intention is non-significant once the cognitive chain is included (with a suppression pattern via trustworthiness). Novelty 2 (experiential boundary): observed valence moderates (negative outcomes further reduce intention), and prominence (hallucination + negative consequence) intensifies the detriment; explains why main effects may appear muted if post-use experience is not captured. Methodology with high ecological validity (real outputs + consequences).	Broaden to heterogeneous samples and real-world contexts (observed behaviours); explore cultural moderators and repeated exposure; test mitigation strategies (RAG, disclaimers, 'show evidence', conservative abstention) and recovery protocols when negative consequences occur.