

Booking Platform, Generative AI, or Open Web? An Experimental Comparison of User Experience in Tourism Accommodation Search

Antun Biloš

Josip Juraj Strossmayer University
of Osijek

Faculty of Economics and
Business

Trg Ljudevita Gaja 7
31000 Osijek, Croatia
antun.bilos@efos.hr

Berislav Andrić

Josip Juraj Strossmayer University
of Osijek

Faculty of Tourism and Rural
Development

Vukovarska 17,
34000 Požega, Croatia
bandrlic@ftrr.hr

Vladimir Ristanović

Institute of European Studies
Trg Nikole Pašića 11

Belgrade 11000, Serbia
vladimir.ristanovic@ies.rs

Abstract. *The introduction of generative AI tools has disrupted established patterns of accommodation search, where online travel agencies have long held a dominant position. This study compares the user experience of three search conditions in a controlled experiment with 165 participants across Croatia, Bosnia and Herzegovina, and Serbia. Booking.com achieved the highest pragmatic UX, ChatGPT showed the highest hedonic UX descriptively, though this difference did not reach statistical significance. Decision time did not differ between conditions. Decision-making experience, operationalized as age within the student sample, moderated these effects: older participants reported markedly better experience with AI and worse experience with open-web search, while their evaluation of Booking.com remained stable.*

Keywords. generative AI, user experience, UEQ-S, online travel agencies, ChatGPT, Booking.com, accommodation search, experimental study, decision-making experience, technology adoption

1 Introduction

Online travel agencies (OTAs) such as Booking.com, Airbnb, and Expedia have established themselves as dominant intermediaries in the European hotel sector, where Booking.com alone has captured a 71% share of the European OTA market and continues to expand its dominance (HOTREC, 2024; Fernández et al., 2020; Kalinić et al., 2019). These platforms have refined their interfaces over decades, offering structured search, aggregated reviews, comparison tools, and integrated booking flows that have become the default reference for how consumers expect digital accommodation search to work.

The introduction of ChatGPT in late 2022 disrupted this standard. A generative AI interface based on natural-language dialogue offers an alternative paradigm that does not rely on filters, lists, or photographs, but instead synthesizes information through conversation (Mich & Garigiano, 2023). Early empirical work has documented growing user willingness to consult ChatGPT for travel planning tasks (Biloš & Budimir, 2024; Arora et al., 2025), while controlled comparisons with established alternatives have begun to appear (Luo et al., 2025; De Cicco et al., 2025; Nie et al., 2025). The result is a conditional rather than uniform AI advantage, with effects depending on task type, user characteristics, and design context.

Despite this growing interest, the user experience of generative AI in tourism search has not yet been systematically compared with both major OTA platforms and unstructured open-web search using standardized UX instruments. Standardized UX evaluations of western OTAs such as Booking.com remain limited, and AI-based travel assistants represent an emerging domain with almost no standardized UX evaluations to date. The present study addresses this gap through a controlled experimental comparison of three accommodation search conditions, evaluated using the short version of the User Experience Questionnaire (UEQ-S; Schrepp et al., 2017) alongside behavioral and self-report outcomes. To the authors' knowledge, this is among the first experimental studies to compare an established booking platform, a generative AI interface, and restricted open-web search within the same accommodation-search task using a standardized UX instrument.

Beyond the descriptive comparison, the study examines a moderation mechanism supported by empirical evidence in related research but not yet tested experimentally in this domain. Recent work has shown that experience moderates the predictors of AI-tool

adoption intention; less experienced users rely more on habit and social influence, whereas more experienced users develop more differentiated evaluations (Biloš et al., 2025). Building directly on this finding, the present study extends the moderation logic from self-reported intention to actual user experience in a realistic decision-making task. The hypothesis under test is that decision-making experience, operationalized as age within a student population, reshapes how users evaluate generative AI relative to established alternatives.

The study was conducted with 165 participants across three countries (Croatia, Bosnia and Herzegovina, Serbia) using a standardized accommodation search scenario in Graz, Austria. The remainder of the paper is organized as follows. Section 2 develops the theoretical background and presents the hypotheses. Section 3 describes the experimental design, instrument, sample, and analytical strategy. Section 4 reports the empirical findings, including the moderation analysis and qualitative illustrations. Section 5 discusses theoretical and practical implications and outlines limitations and directions for future research.

2 Theoretical Background

2.1 User Experience and the UEQ-S Framework

User experience (UX) is defined by ISO 9241-210 (2019) as a person's perceptions and responses resulting from the use or anticipated use of a product, system, or service. This conceptualization deliberately spans pragmatic dimensions related to instrumental task performance and hedonic dimensions related to subjective enjoyment and emotional engagement. The User Experience Questionnaire and its short version (UEQ-S) operationalize this distinction through eight semantic differential items that yield two scale scores, Pragmatic Quality and Hedonic Quality, alongside an aggregate score that summarizes overall UX (Schrepp et al., 2017). Recent cross-cultural adaptations of the instrument confirm its applicability across linguistic contexts when the underlying factor structure is preserved (Melissourgous et al., 2025; Juric et al., 2024).

The pragmatic-hedonic distinction is theoretically valuable for comparing digital tools because the two dimensions can be dissociated. A tool may be pragmatically efficient yet hedonically dull, or hedonically engaging yet pragmatically frustrating. Standardized instruments such as UEQ-S have been increasingly applied to consumer-facing digital interfaces (Islami et al., 2023; Jung et al., 2026), although standardized UX evaluations of western online travel agencies and AI-based travel assistants remain limited.

2.2 Generative AI and Established Tools in Consumer Decision-Making

Travel decisions have long been mediated by online channels. Online travel agencies (OTAs) have established a dominant position in accommodation distribution, accounting for over 90% of online hotel reservations in some markets, with Booking.com leading the segment (Fernández et al., 2020; Kalinić et al., 2019). Empirical research on consumer behavior within OTA platforms has identified price, ratings, search position, and informational completeness as key drivers of conversion across the search and booking funnel (Xie et al., 2019; Pinto et al., 2019). The instrumental quality of OTA interfaces, including filters, photographs, and aggregated reviews, has been refined over decades of iterative design.

The introduction of ChatGPT in late 2022 changed the landscape by offering an alternative interaction paradigm based on natural-language dialogue rather than structured filters and lists. Conceptual work mapped the technological potential of large language models for tourism early on (Mich & Garigiano, 2023), and a rapidly expanding empirical literature has begun to examine their actual use. Survey-based research grounded in extended versions of the Technology Acceptance Model (Davis, 1989) and UTAUT2 (Venkatesh, Thong, & Xu, 2012) has identified habit, performance expectancy, and hedonic motivation as the strongest predictors of behavioral intention to use ChatGPT among Generation Z users (Biloš & Budimir, 2024; Arora et al., 2025). Performance expectancy and effort expectancy show inconsistent effects across this literature, suggesting that ease of use is no longer a meaningful differentiator between AI and conventional digital tools.

Experimental work has begun to examine how generative AI compares with established alternatives. Luo et al. (2025) manipulated search type, travel motive, and AI customization level and found that the relative advantage of generative AI over conventional search engines is conditional rather than uniform. De Cicco et al. (2025) compared AI and human advisors across pre-visit and on-site travel stages and reported stage-dependent preferences. Studies on trust have shown that hallucinations and inaccuracies in AI output reduce perceived accuracy and willingness to follow AI recommendations (Kim et al., 2023; Christensen et al., 2025). At the same time, evidence from age-segmented evaluations of LLM-based travel assistants suggests that acceptability and perceived credibility vary across user groups (Nie et al., 2025). The current literature points toward a tool whose UX value is contingent rather than universal, with the contingency factors only beginning to be mapped.

2.3 Decision-Making Experience as a Moderator

A long tradition in behavioral decision research has examined how accumulated experience moderates decision-makers' use of information aids. Older adults with greater consumer experience adapt their search strategies in complex consumer choice tasks and maintain decision quality through accrued domain knowledge (Queen et al., 2012). Prior domain knowledge similarly reduces age-related differences in web search performance (Sanchiz et al., 2017) and interacts with interface design in three-way patterns where experienced users in supportive interface conditions can match or exceed less experienced peers (Zimmer et al., 2015). The dual-process distinction between automatic and deliberative cognition (Kahneman, 2011) provides a theoretical mechanism: experienced users have richer mental representations of available tools and evaluate them through more discriminating, comparative processing, while less experienced users default to undifferentiated heuristic evaluation.

Generational labels such as "digital natives" have been challenged empirically. Hoffmann, Lutz, and Meckel (2014) showed that web experience and education explain differences in online trust more reliably than birth cohort, and cross-national evidence indicates that technology skill and use are distributed on a continuum rather than dichotomized by generation (Bullen & Morgan, 2011). For research on generative AI specifically, this distinction matters because it implies that observed age effects may reflect accumulated decision-making experience rather than generational technological orientation.

Within technology adoption frameworks, experience operates as a formal moderator (Venkatesh et al., 2012) that reshapes which predictors translate into intention or use. Habit, defined as the extent to which behavior becomes automatic through repeated learning (Limayem, Hirt, & Cheung, 2007), accumulates with experience and has been identified as the single strongest predictor of ChatGPT adoption intention in regional Generation Z populations (Biloš & Budimir, 2024). Building directly on this evidence, Biloš et al. (2025) demonstrated that experience itself moderates the predictors of ChatGPT usage intention: less experienced users rely more heavily on habit and social influence, while more experienced users develop differentiated patterns of evaluation. The mechanism documented at the level of behavioral intention should plausibly extend to the level of actual user experience during real interactions with the tool.

2.4 Research Hypotheses

Based on the literature reviewed above, the following hypotheses guide the analysis:

H1. Tool type affects the pragmatic dimension of user experience, with the established Booking platform

achieving higher UEQ-S Pragmatic scores than open-web search.

H2. Tool type affects the hedonic dimension of user experience, with the generative AI condition achieving higher UEQ-S Hedonic scores than the established Booking platform.

H3. Decision-making experience, operationalized as age within the student population, moderates the effect of tool condition on user experience.

A secondary research question (RQ1) examines whether tool conditions differ in objective and behavioral outcomes, namely decision time, chosen accommodation price, and stated booking intention.

3 Method

3.1 Design and Participants

The study employed a between-subjects experimental design comparing three accommodation search conditions: (E1) booking through the Booking.com platform, (E2) searching with the ChatGPT conversational interface, and (K) open-web search excluding both online travel agencies and large language model tools. The factorial structure additionally included country (Croatia, Bosnia and Herzegovina, Serbia) and gender as design factors with balanced cell sizes targeted in advance.

Participants were recruited from the student populations of several universities in Croatia, Bosnia and Herzegovina, and Serbia. The final analytic sample included 165 participants (49 from Croatia, 59 from Bosnia and Herzegovina, 57 from Serbia), with a balanced gender distribution (83 men, 82 women) and a mean age of 24.7 years ($SD = 5.92$, range 19 to 54). The educational profile reflected the recruitment context, with 55.2% reporting completed secondary education, 22.4% completed undergraduate studies, 16.4% completed graduate studies, and the remaining participants distributed across postgraduate and other categories. Allocation to experimental conditions was implemented at the level of small-group sessions. Participation was voluntary and anonymous, with informed consent obtained on the first page of the survey instrument. The study was conducted in accordance with general research ethics principles for studies involving human participants.

3.2 Procedure and Task

All participants received an identical task scenario presented in their native language. Participants were asked to imagine planning a short trip and to find one accommodation option that they would realistically reserve. Task parameters were fixed across all conditions to control for choice complexity: destination Graz (Austria), two adults, two nights from Friday 17 April to Sunday 19 April 2026, maximum total budget of 300 EUR for both nights, location

preferably within 3 km of Hauptplatz, and accommodation type restricted to hotel rooms, private rooms, studios, or apartments (excluding hostel dormitories). Participants were instructed not to complete an actual reservation; the task ended when they indicated their final choice. The protocol communicated a guideline of approximately 15 minutes for task completion, which was not strictly enforced. In practice, 98.8% of participants completed the task within the 15-minute window. The full sample showed a mean decision time of 215.4 seconds ($Mdn = 177.0$ s, range 21 to 986 s).

The experimental manipulation operated through tool restriction. Participants in E1 used Booking.com exclusively. Participants in E2 used ChatGPT exclusively, remaining within the conversational interface and not opening external websites in separate tabs. Participants in K used open-web search through a search engine and direct websites of accommodation providers or tourism organizations, while explicitly avoiding online travel agencies and large language model interfaces. A minimum decision time of 20 seconds was established as a data quality criterion based on observations from pilot testing, where shorter completions consistently reflected disengaged or hurried responding rather than genuine task performance. Responses below this threshold were excluded prior to analysis. Protocol adherence was monitored through a separate deviations log completed by the researcher after each session, designed to preserve the anonymity of participant responses while documenting any procedural irregularities. The number of recorded deviations was small and did not show a systematic pattern across conditions. The analytic sample of 165 participants reflects the dataset following these data quality checks.

3.3 Measures

Decision time was recorded automatically by the survey platform (Alchemer) as the duration spent on the task page, from the moment the task was displayed until participants clicked the corresponding button to indicate they had made their decision.

User experience was measured using the short version of the User Experience Questionnaire (UEQ-S; Schrepp et al., 2017). The UEQ-S comprises eight items in semantic differential format on a seven-point scale, organized into two dimensions of four items each: Pragmatic Quality (obstructive/supportive, complicated/easy, inefficient/efficient, confusing/clear) and Hedonic Quality (boring/exciting, not interesting/interesting, conventional/inventive, usual/leading-edge). Following standard UEQ practice, raw responses were transformed to the official metric ranging from -3 to +3 by subtracting 4 from each item score, with values above +0.8 indicating positive evaluations and values below -0.8 indicating negative evaluations. The instrument was administered in the language version corresponding to each

participant's country of residence. The Croatian items followed the order of the original UEQ-S list with adjustments for cross-country comparability with Bosnian and Serbian versions, departing in some pair formulations from the validated Croatian UEQ version (Juric et al., 2024). Cross-cultural adaptation of UEQ-S for related linguistic contexts has recently been reported by Melissourgous et al. (2025), supporting the legitimacy of regional adaptations when the underlying factor structure is preserved.

Behavioral and evaluative outcomes following the task included: total price of the chosen accommodation in euros (open numeric entry), accommodation type, booking intention measured as the perceived likelihood of actually reserving the chosen accommodation (1 = not at all likely, 7 = extremely likely), perceived task ease (1 = very difficult, 7 = very easy), and prior familiarity with the destination (1 = not familiar at all, 5 = very familiar). An open-ended question invited participants to comment on what helped or hindered their decision.

3.4 Analytical Strategy

Statistical analyses were conducted using SPSS 29.0.2.0 and Microsoft Excel. Analyses proceeded in three stages. First, between-condition differences on the seven primary outcome variables were tested using one-way ANOVA, with effect sizes reported as omega squared (ω^2) to provide unbiased estimates given unequal cell sizes. Second, the central hypothesis on the moderating role of decision-making experience was tested using hierarchical linear regression, with age entered as a continuous moderator interacting with experimental condition (dummy-coded with K as reference). Median split visualizations are used for illustrative purposes only, while the statistical inference relies on the continuous specification. Third, robustness across the three countries was examined by re-estimating the central interaction model with country dummies and country $\times$ condition terms, following recent recommendations in cross-cultural OTA research (Filieri et al., 2021). Open-ended responses (v09) were thematically reviewed to identify illustrative quotations supporting the quantitative findings. Levene's test indicated unequal group variances for two outcomes (task ease and UEQ-S Pragmatic Quality); Welch's robust ANOVA yielded equivalent conclusions for these variables (Welch's $F = 15.06$ and 5.88 , both $p < .01$), and standard ANOVA results are reported throughout.

Generative AI tools (Claude Opus 4.7, ChatGPT 5.5, Gemini 3.1 Pro) and the literature discovery tool SciSpace were consulted at various stages for support tasks such as study design refinement, literature mapping, and drafting iterations of the manuscript. Grammarly was used for language editing. All scientific decisions, statistical analyses, interpretations, and final wording remained the responsibility of the authors.

4 Results

4.1 Main Effects of Tool Type

Table 1 reports descriptive statistics and one-way ANOVA results for the seven primary outcome variables across the three experimental conditions.

Table 1. Means and one-way ANOVA results across tool conditions

Variable	Booking	ChatGPT	Open-web	F (p)	ω^2
Decision time (s)	211.3 (197.6)	212.8 (133.6)	223.4 (173.5)	0.08 (.925)	< .001
Price (€)	203.7 (62.4)	214.4 (52.0)	217.7 (51.7)	1.00 (.370)	< .001
Booking intention (1–7)	5.52 (1.16)	5.15 (1.22)	4.80 (1.48)	4.59 (.012)	.042
Task ease (1–7)	6.40 (0.91)	5.90 (1.45)	5.28 (1.23)	12.91 (.001)	.126
UEQ-S Pragmatic (–3 to +3)	1.84 (1.07)	1.61 (1.42)	1.03 (1.38)	5.90 (.003)	.056
UEQ-S Hedonic (–3 to +3)	0.76 (1.27)	1.22 (0.92)	0.77 (1.03)	2.99 (.053)	.024
UEQ-S Overall (–3 to +3)	1.30 (1.02)	1.42 (1.01)	0.90 (1.08)	3.51 (.032)	.030

Note. Booking $n = 67$, ChatGPT $n = 48$, Open-web $n = 50$. Means with standard deviations in parentheses. F values report $F(2, 162)$. Effect sizes (ω^2) reported as $<.001$ indicate negligible effects.

The conditions differed substantially on subjective task ease ($\omega^2 = .126$) and on UEQ-S Pragmatic Quality ($\omega^2 = .056$). Booking achieved the highest scores on both measures, with Open-web search performing the weakest (Fig. 1). Booking intention followed the same ordering, although with a smaller effect size ($\omega^2 = .042$). The pattern on UEQ-S Hedonic Quality reversed the ordering on the pragmatic dimension, with ChatGPT scoring highest ($M = 1.22$) and Booking lowest ($M = 0.76$). On the composite UEQ-S Overall score, ChatGPT slightly outperformed Booking, with Open-web again being the lowest. Tukey HSD post-hoc tests confirmed that Booking significantly exceeded Open-web on UEQ-S Pragmatic Quality (mean difference = 0.81, $p = .003$). While ChatGPT showed a higher mean on UEQ-S Hedonic Quality than Booking (mean difference = 0.47), this difference did not reach statistical significance (Tukey HSD $p = .069$).

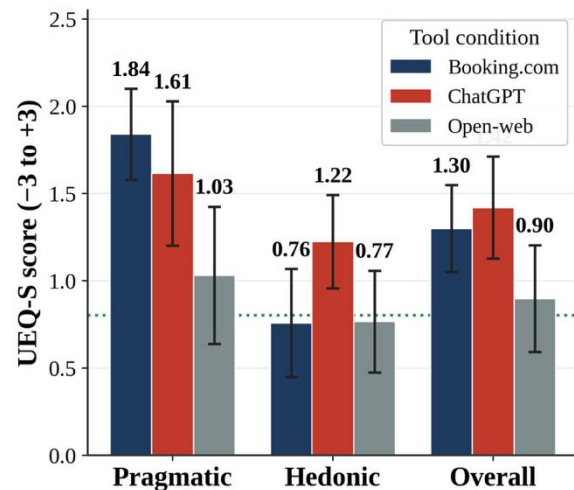


Figure 1. UEQ-S Pragmatic, Hedonic, and Overall scores across tool conditions, with 95% confidence intervals. The dotted reference line marks the +0.8 benchmark for positive evaluations.

Two outcomes showed no statistically reliable variation between tools. The price of the chosen accommodation was similar across conditions, with means ranging from 204 to 218 EUR, well below the 300 EUR task limit. Decision time was likewise indistinguishable between tools (means ranging from 211 to 223 seconds), with ω^2 approaching zero. Given the right-skewed distribution of raw decision time, the analysis was repeated on log-transformed values; the conclusion remained unchanged ($F(2,162) = 0.25$, $p = .781$).

4.2 The Fluency-Time Disconnect

The combined pattern in Table 2 reveals a notable disconnect between objective and subjective task indicators (Fig. 2). Although participants spent virtually identical amounts of time on the task across conditions, their subjective evaluations of task ease differed by more than 1 full-scale point on a 7-point scale (Booking: 6.40; ChatGPT: 5.90; Open-web: 5.28). The within-sample correlation between decision time and perceived task ease was negligible ($r = -.08$, $p = .288$), confirming that the time spent on the task carried little information about how the experience felt to participants. Tool conditions did not differ in how long they took, but they clearly differed in how easy and pragmatically well-functioning they were perceived to be. This disconnect motivates the central question that follows: if subjective experience varies independently of objective effort, what individual-level factors shape the magnitude and direction of these subjective evaluations?

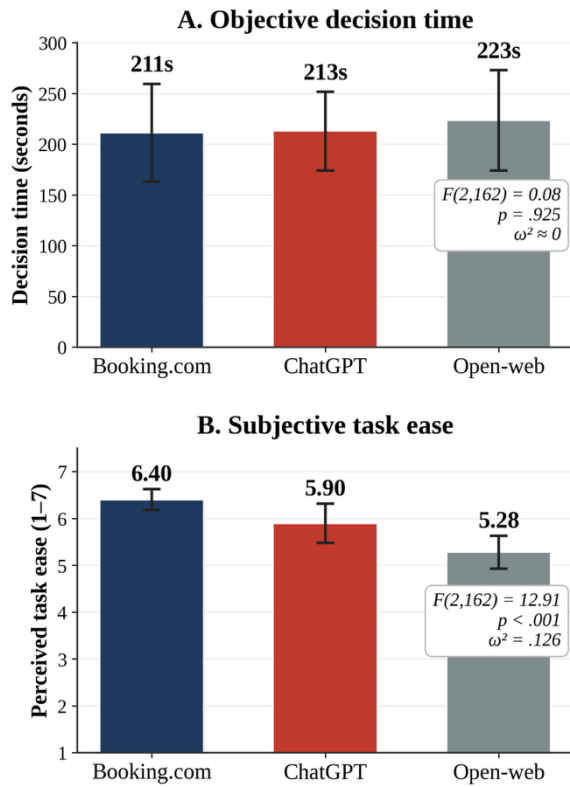


Figure 2. The fluency-time disconnect. Panel A shows decision time in seconds. Panel B shows perceived task ease on a 1–7 scale. Error bars represent 95% confidence intervals.

4.3 Decision-Making Experience as a Moderator

The central hypothesis predicted that decision-making experience would moderate the effect of tool condition on user experience. Hierarchical regression analyses tested this prediction on the two UEQ-S outcomes that showed reliable between-tool differences: Pragmatic Quality and Overall UX. Age was used as a continuous proxy for accumulated decision-making experience and was mean-centered before model estimation. Tool conditions were dummy-coded with Open-web (K) as the reference category. Gender was retained as a covariate. Because Open-web was used as the reference condition, the age coefficient represents the age slope in the restricted open-web condition, while the interaction terms indicate how much the Booking and ChatGPT age slopes differ from that reference slope. Thus, a positive ChatGPT $\times$ Age interaction means that the association between age and UX was more positive in the ChatGPT condition than in the restricted open-web search.

Table 2. Hierarchical regression predicting UEQ-S Pragmatic Quality and UEQ-S Overall

Step / Predictor	B	SE	t	B	SE	t
	UEQ-S Pragmatic			UEQ-S Overall		
Step 1: Controls						
Age (centered)	.02	.02	0.92	.01	.01	0.79
Gender (Male=1)	-.56	.21	-2.73**	-.29	.17	-1.76
R ²	.045			.020		
Step 2: Main effects (ΔR ²)	.063 **			.041 *		
Booking (vs Open- web)	.78	.24	3.30**	.38	.19	1.98*
ChatGPT (vs Open- web)	.58	.26	2.27*	.53	.21	2.53*
Step 3: Tool × Age (ΔR ²)	.087 ***			.076 **		
Booking × Age	.12	.04	2.99**	.09	.03	2.64**
ChatGPT × Age	.21	.05	4.06***	.16	.04	3.67***

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

The interaction terms made a substantial contribution to the model in both cases. For UEQ-S Pragmatic, adding the two Tool $\times$ Age interactions increased explained variance from 10.8% to 19.5%, with both interaction terms reaching statistical significance, and the ChatGPT $\times$ Age slope showing the strongest effect in the model ($B = .21$, $t = 4.06$, $p < .001$). The pattern was parallel for UEQ-S Overall (final $R^2 = .137$).

To interpret the interactions, simple slopes of age were estimated within each tool condition, controlling for gender (Fig. 3). In the Booking condition, age was unrelated to UEQ-S Pragmatic ($B = +.02$, $p = .207$). In the ChatGPT condition, age showed a strong positive slope ($B = +.12$, $t = 2.76$, $p = .008$), indicating that older participants reported markedly better pragmatic UX than younger ones. In the Open-web condition, the slope reversed sign and reached statistical significance in the opposite direction ($B = -.10$, $t = -2.48$, $p = .017$), with older participants reporting worse pragmatic UX than younger ones. The same crossover pattern held for UEQ-S Overall (Booking $B = +.02$, $p = .256$; ChatGPT $B = +.09$, $p = .009$; Open-web $B = -.07$, $p = .037$).

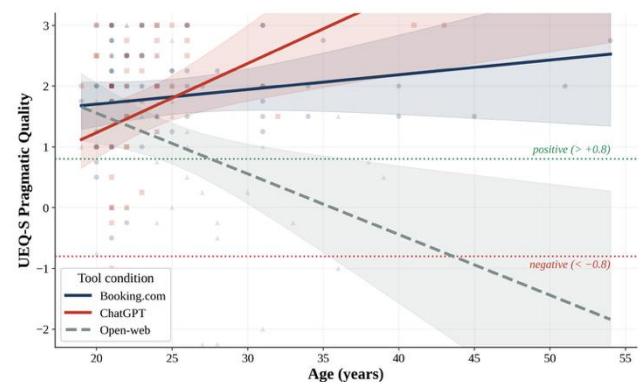


Figure 3. Tool $\times$ Age interaction on UEQ-S Pragmatic Quality. Lines represent regression-

predicted scores at the sample mean for gender, with shaded bands indicating 95% confidence intervals.

In summary, Booking provided a stable user experience across the age range observed in the sample, with younger and older participants converging on the same evaluation. ChatGPT and Open-web search behaved in the opposite manner: as accumulated decision-making experience increased, ChatGPT was evaluated more favorably while Open-web was evaluated less favorably. The effect of tool condition on user experience depended systematically on participant experience rather than operating uniformly across the sample.

4.4 Robustness Across Countries

To assess the generality of the moderation pattern, the Tool $\times$ Age regression was re-estimated separately for each national subsample. A statistically significant Tool $\times$ Age interaction was reproduced in each country, although the specific term reaching significance varied: ChatGPT $\times$ Age in Croatia ($B = +0.24$, $p = .011$) and Serbia ($B = +0.73$, $p = .007$), and Booking $\times$ Age in Bosnia and Herzegovina ($B = +0.12$, $p = .030$). Cell sizes ranged from 14 to 28 participants per condition $\times$ country, limiting statistical power for individual subsample tests, and the magnitude of the Serbian coefficient should be interpreted with caution given the small subsample size. The cross-country replication of the directional pattern is consistent with the pooled finding.

4.5 Qualitative Illustrations

Open-ended responses were thematically reviewed to identify quotations that illustrated the quantitative pattern (quotations were translated into English by the authors). Three themes emerged that map directly onto the regression findings.

First, participants in the Open-web condition described systematic difficulty avoiding the gravitational pull of the Booking platform, even when explicitly instructed to do so. Comments such as "Booking kept getting in my way" (Serbia, age 23) and "I had Booking offers showing up although I was not allowed to use them" (Serbia, age 25) suggest that the Open-web condition was not simply less efficient but actively effortful, which is consistent with its lowest pragmatic UX score and the negative age slope on this outcome.

Second, the youngest participants in the ChatGPT condition expressed explicit distrust of generative AI. Statements such as "I always doubt the results of AI tools and would rather check accommodations manually" (Bosnia and Herzegovina, age 22) and "The accuracy of LLM information is questionable to me in general" (Croatia, age 21) align with the lower UX evaluations among younger participants in this condition. By contrast, slightly older participants in the

same condition described positive engagement with the conversational interface, including formulations such as "ChatGPT was a piece of cake to use, that is how I will look for accommodation in the future" (Serbia, age 24).

Third, one participant explicitly noted that ChatGPT proposed only a narrow subset of accommodation types, which echoes the descriptive observation that hotels dominated the chosen options in the ChatGPT condition. This kind of constraint on the choice space, although not the focus of the present analysis, supports the broader interpretation that each tool reshapes not only how users feel about the search but also what they actually consider.

5 Discussion

5.1 Synthesis of Findings

Three findings stand out from the experimental comparison. The pragmatic dimension of user experience varied substantially across tools, with Booking achieving the highest scores and open-web search the lowest, while ChatGPT fell between the two. The hedonic dimension reversed this ordering, with ChatGPT scoring highest. Decision time showed virtually no variation between tools despite large differences in subjective task ease, which produced a near-zero correlation between objective time spent and perceived ease. Most centrally for the contribution of this paper, decision-making experience moderated the effect of tool condition on UX in the predicted direction. Older participants reported markedly better pragmatic UX with ChatGPT and markedly worse pragmatic UX with open-web search than younger participants, while their evaluation of Booking remained stable across the age range observed in the sample. The same crossover pattern was reproduced in UEQ-S Overall and maintained the same direction across the three national subsamples. Taken together, these findings provide clear support for H1 and H3. H2 was not supported. Although ChatGPT showed a higher mean on UEQ-S Hedonic Quality than Booking.com, the difference did not reach statistical significance (Tukey HSD $p = .069$). This suggests that the hedonic appeal of generative AI should not be overstated. In this task context, accommodation search remained strongly instrumental, and participants may have evaluated enjoyment through the lens of decision confidence rather than novelty alone.

5.2 Theoretical Contributions

The moderation finding extends the recent demonstration that experience moderates the predictors of generative AI adoption intention (Biloš et al., 2025) into the domain of actual user experience. The earlier survey-based work showed that less experienced users rely more heavily on habit and social influence in

forming their adoption intentions. The present experimental results indicate a parallel pattern in real interaction: less experienced users showed relatively flat UX evaluations across tool conditions, whereas more experienced users discriminated sharply between AI, the established OTA, and unstructured open-web search. The convergence of self-reported intention and behavioral UX evidence strengthens the case for experience as a fundamental moderator of how users translate AI capabilities into perceived value. One possible explanation is that relatively older participants within the student sample benefited more from ChatGPT because the conversational interface reduced the need to manage fragmented search results, compare multiple sources manually, and avoid dominant OTA links. In contrast, restricted open-web search may have imposed a greater coordination burden, especially because participants had to search the web while deliberately avoiding the platforms that typically structure accommodation searches.

The fluency-time disconnect carries both methodological and substantive implications for UX research. Tools that produced very different subjective experiences required statistically indistinguishable amounts of time to complete the same task. Inferring tool efficiency from objective time measurements, or inferring user experience from objective time savings, would have produced misleading conclusions in this dataset. The result is consistent with broader work on processing fluency, which has shown that subjective ease can be experimentally dissociated from objective performance. For comparative evaluations of AI and conventional digital tools, this dissociation suggests that subjective UX instruments and objective performance measures address complementary rather than overlapping questions.

A third contribution concerns the position of generative AI in the existing ecosystem. The pattern of pragmatic-hedonic dissociation observed in our data is consistent with prior UTAUT2 evidence (Biloš & Budimir, 2024): hedonic motivation has been identified as a key driver of ChatGPT adoption, although our test of the corresponding hedonic UX advantage in actual interaction did not reach statistical significance, the descriptive pattern was directionally consistent with this prediction. At the same time, established OTA platforms retain a clear pragmatic advantage rooted in decades of refinement, including filters, photos, and reviews. The empirical question raised by Mich and Garigiano (2023) regarding whether AI will replace OTA platforms therefore receives a nuanced answer in the present data: it depends on which UX dimension is at stake and on which segment of users is considered. The results therefore suggest that Booking.com and ChatGPT should not be interpreted as direct substitutes on a single UX continuum: Booking.com currently offers stronger pragmatic assurance, while ChatGPT offers a more engaging but not statistically superior hedonic experience.

5.3 Practical Implications

For tourism platforms, the findings suggest two priorities. First, OTA platforms retain a clear advantage on pragmatic dimensions and should defend this advantage by continuing to invest in structured affordances rather than imitating AI-style interaction. Second, the segment of users for whom generative AI is most attractive is not the youngest population but those with somewhat greater accumulated decision-making experience. Marketing and integration strategies for AI-augmented search should reflect this segmentation rather than assume uniform appeal across age cohorts. For AI tool designers, the disconnect between subjective ease and objective time indicates that hedonic engagement does not automatically translate into productivity gains, and that pragmatic features such as visual content, link validity, and filtering remain necessary complements to conversational fluency.

5.4 Limitations

Several limitations should be kept in mind when interpreting these results. The sample was drawn from student populations across three countries, and the age range observed (19 to 54 years, with most participants under 30) means that the observed moderation describes the transition from undergraduate to early professional life rather than a full generational comparison. Therefore, the findings should not be generalized to the full population of accommodation bookers, older tourists, or less digitally active users. Rather, they describe how relatively young and educated users evaluate competing digital search tools under controlled task conditions. The study used a single destination, a single accommodation task, and a single price ceiling, which may limit the external validity of specific descriptive findings such as the dominance of hotels in chosen accommodations. The cross-sectional design does not allow conclusions about how repeated use would change the observed UX patterns. ChatGPT was used in a single configuration, and rapid iteration of generative AI tools means that specific findings may date faster than the underlying theoretical pattern.

5.5 Future Research

Three directions appear particularly promising. First, replication on non-student populations with broader age ranges would test whether the moderation pattern generalizes beyond the early professional segment. Second, longitudinal designs that track UX over repeated use would show whether the dissociation between objective time and subjective experience persists or diminishes as users build mental models of the tools. Third, comparative experiments with additional generative AI interfaces (Gemini, Claude, Perplexity) and different tourism task types (route

planning, local activities, multi-stop itineraries) would clarify whether the patterns observed for ChatGPT in accommodation search reflect properties of the tool category or specific design choices of one platform.

6 Conclusion

The user experience of generative AI in accommodation search is neither uniformly superior nor inferior to established alternatives, but rather depends on the UX dimension examined and the user's prior decision-making experience. ChatGPT showed a directional advantage on hedonic UX, while Booking.com retains a clear pragmatic advantage built over decades of interface refinement. The moderation pattern suggests that AI integration in tourism platforms should consider user segmentation rather than assume uniform appeal. The dissociation between perceived ease and objective time spent further indicates that comparative evaluations of AI and conventional digital tools require both UX instruments and behavioral measures, since the two address complementary questions.

Acknowledgments

This research is funded by the Croatian Science Foundation (HRZZ) research project "Tourism 5.0: Customer Satisfaction through the Digital Transformation of Tourist Destinations", grant no. IP-2025-02-1899 (Croatian Science Foundation).

References

- Arora, N., Manchanda, P., Aggarwal, A., & Maggo, V. (2025). Tapping generative AI capabilities: A study to examine continued intention to use ChatGPT in the travel planning. *Asia Pacific Journal of Tourism Research*, 30(10), 1303–1322. <https://doi.org/10.1080/10941665.2024.2405134>
- Biloš, A., & Budimir, B. (2024). Understanding the adoption dynamics of ChatGPT among Generation Z: Insights from a modified UTAUT2 model. *Journal of Theoretical and Applied Electronic Commerce Research*, 19(2), 863–879. <https://doi.org/10.3390/jtaer19020045>
- Biloš, A., Turkalj, D., & Kelić, I. (2025). The role of experience in AI adoption: The moderating effect on predictors of ChatGPT usage intention among Generation Z. In *Proceedings of the 9th FEB International Scientific Conference* (pp. 289–298). University of Maribor Press. <https://doi.org/10.18690/um.epf.5.2025.27>
- Bullen, M., & Morgan, T. (2011). Digital learners not digital natives. *La Cuestión Universitaria*, 7, 60–68.
- Christensen, J., Hansen, J. M., & Wilson, P. (2025). 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>
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340.
- De Cicco, R., Francioni, B., Dini, M., Curina, I., Filieri, R., & Cioppi, M. (2025). Generative AI and human advice: User perceptions, intentions, and behavior across pre-visit and on-site travel stages. *Journal of Travel Research*. <https://doi.org/10.1177/00472875251364222>
- Fernández, J. M., Vall-Llosera, L., & Moya, D. H. (2020). Analysis of OTA impact on hotel reservations: Case study: Bogotá. *Turismo y Sociedad*, 27, 145–159. <https://doi.org/10.18601/01207555.n27.08>
- Filieri, R., & Mariani, M. (2021). The role of cultural values in consumers' evaluation of online review helpfulness: A big data approach. *International Marketing Review*, 38(6), 1267–1288. <https://doi.org/10.1108/IMR-07-2020-0172>
- HOTREC. (2024). European hotel distribution study 2024: Results for the reference year 2023 (R. Schegg, Author). HOTREC & HES-SO Valais-Wallis. https://www.hotrec.eu/media/static/files/import/all_news_2024_2024_21/hotrec-distribution-study-2024.pdf
- Hoffmann, C. P., Lutz, C., & Meckel, M. (2014). Digital natives or digital immigrants? The impact of user characteristics on online trust. *Journal of Management Information Systems*, 31(3), 138–171.
- ISO. (2019). ISO 9241-210:2019 Ergonomics of human-system interaction. Part 210: Human-centred design for interactive systems. International Organization for Standardization.
- Islami, R., Hilabi, S. S., & Hananto, A. (2023). Analisis user experience aplikasi Traveloka dan Tiket.Com menggunakan metode User Experience Quesionnaire. REMIK: Riset dan E-Jurnal Manajemen Informatika Komputer, 7(1), 497–505. <https://doi.org/10.33395/remik.v7i1.12106>
- Jung, E., Shi, S., & Rew, D. (2026). Unlocking trust: How UX builds trust and engagement in online travel agencies. *Tourism and Hospitality Research*. <https://doi.org/10.1177/14673584261418145>
- Juric, M., Pehar, F., & Peša Pavlović, N. (2024). Translation, psychometric evaluation, and

- validation of the Croatian version of the User Experience Questionnaire (UEQ). *International Journal of Human-Computer Interaction*, 40(7), 1644–1657.
<https://doi.org/10.1080/10447318.2022.2147713>
- Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Kalinić, Z., & Novaković, M. (2019). Online accommodation booking habits and attitudes of Serbian travelers. *Menadžment u Hotelijerstvu i Turizmu*, 7(2), 11–23.
<https://doi.org/10.5937/menhotur1902011K>
- Kim, J. H., Kim, J., Park, J., Kim, C., & Jhang, J. (2023). When ChatGPT gives incorrect answers: The impact of inaccurate information by generative AI on tourism decision-making. *Journal of Travel Research*.
<https://doi.org/10.1177/00472875231212996>
- Limayem, M., Hirt, S. G., & Cheung, C. M. K. (2007). How habit limits the predictive power of intention: The case of information systems continuance. *MIS Quarterly*, 31(4), 705–737.
- Luo, X., Xu, D., Li, Y., & Wan, L. C. (2025). Advancing information search through GenAI: The roles of search type, travel motive and GenAI customization level. *International Journal of Contemporary Hospitality Management*, 37(5), 1725–1743. <https://doi.org/10.1108/IJCHM-06-2024-0941>
- Melissourgos, G., & Katsanos, C. (2025). UEQ-GR and UEQ-S-GR: Towards a Greek adaptation of the User Experience Questionnaire and its short version. In *CHI Greece '25: Proceedings of the 3rd International Conference of the ACM Greek SIGCHI Chapter* (203–211).
<https://doi.org/10.1145/3749012.3749059>
- Mich, L., & Garigliano, R. (2023). ChatGPT for e-tourism: A technological perspective. *Information Technology & Tourism*, 25, 1–12.
- Nie, F., Yuhan, S., Hou, T., Shi, S., & Chen, X. (2025). Enhancing tourist experience across age groups with large language models: Evaluating ChatGPT as a virtual tour guide. *Proceedings of the ACM on Smart Tourism*.
<https://doi.org/10.1145/3745238.3745457>
- Pinto, P., Castro, C., & Mota, A. (2019). Online travel agencies: Factors influencing tourists' purchase decisions. *Tourism & Management Studies*, 15(2), 7–20.
- Queen, T. L., Hess, T. M., Ennis, G. E., Dowd, K., & Grün, D. (2012). Information search and decision making: Effects of age and complexity on strategy use. *Psychology and Aging*, 27(4), 817–824.
- Sanchiz, M., Chin, J., Chevalier, A., Fu, W.-T., Amadiéu, F., & He, J. (2017). Searching for information on the web: Impact of cognitive aging, prior domain knowledge and complexity of the search problems. *Information Processing and Management*, 53(1), 281–294.
<https://doi.org/10.1016/j.ipm.2016.09.003>
- Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017). Design and evaluation of a short version of the User Experience Questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(6), 103–108.
- Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, 36(1), 157–178.
- Xie, K. L., & Lee, Y. J. (2020). Hotels at fingertips: Informational cues in consumer conversion from search, click-through, to book. *Journal of Hospitality and Tourism Technology*, 11(1), 49–67. <https://doi.org/10.1108/JHTT-03-2017-0026>
- Zimmer, J. C., Tams, S., Craig, K., Thatcher, J., & Pak, R. (2015). The role of user age in task performance: Examining curvilinear and interaction effects of user age, expertise, and interface design. *Journal of Business Economics*, 85(4), 367–394.