University of Reading

Generative AI in digital engagement: a quasi-experimental study of tourist sentiment

Article

Published Version

Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Open Access

Ali, W. ORCID: https://orcid.org/0000-0002-6200-8580, Kasturiratne, D. ORCID: https://orcid.org/0009-0000-1959-0918, Ameer, I. ORCID: https://orcid.org/0000-0002-2379-7863 and Bhaskar, S. ORCID: https://orcid.org/0000-0002-8835-1918 (2026) Generative AI in digital engagement: a quasi-experimental study of tourist sentiment. The Service Industries Journal. ISSN 1743-9507 doi: 10.1080/02642069.2026.2638226 Available at https://centaur.reading.ac.uk/128929/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1080/02642069.2026.2638226

Publisher: Routledge

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

University of Reading

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

The Service Industries Journal

ISSN: 0264-2069 (Print) 1743-9507 (Online) Journal homepage: www.tandfonline.com/journals/fsij20

Generative AI in digital engagement: a quasi-experimental study of tourist sentiment

Waqas Ali, Dulekha Kasturiratne, Irfan Ameer & Shikhar Bhaskar

To cite this article: Waqas Ali, Dulekha Kasturiratne, Irfan Ameer & Shikhar Bhaskar (06 Mar 2026): Generative AI in digital engagement: a quasi-experimental study of tourist sentiment, The Service Industries Journal, DOI: 10.1080/02642069.2026.2638226

To link to this article: https://doi.org/10.1080/02642069.2026.2638226

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 06 Mar 2026.

Submit your article to this journal

Article views: 916

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=fsij20

THE SERVICE INDUSTRIES JOURNAL
https://doi.org/10.1080/02642069.2026.2638226

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Generative AI in digital engagement: a quasi-experimental study of tourist sentiment

生成式人工智能在数字互动中的应用研究：基于游客情感的准实验分析

Waqas Ali [a], Dulekha Kasturiratne [b], Irfan Ameer [a,c] and Shikhar Bhaskar [a]

[a]Plymouth Business School, University of Plymouth, Plymouth, UK; [b]Henley Business School, University of Reading, Reading, UK; [c]Turku University of Applied Sciences, Turku, Finland

ABSTRACT

This study examines whether generative AI can enhance tourist sentiment in online reviews by acting as a consistent and scalable form of digital engagement. Using a quasi-experimental Difference-in-Differences design, we analysed 11,393 reviews collected for six months from five Indian restaurants in a UK tourist city, where one restaurant adopted ChatGPT-generated responses and four served as controls. Results show that AI mediated responses produced a significant improvement in tourist sentiment, with similar effects across both social (Google, Facebook) and delivery-oriented platforms (Deliveroo, Just Eat, Uber Eats). Response features such as tone, personalisation, and length had limited additional influence, indicating that the presence of a response matters more than its specific stylistic qualities. The findings suggest that GAI-mediated responses influence tourist sentiment primarily by signalling organisational attentiveness and relational legitimacy, rather than through nuanced stylistic features of the response. The study demonstrates how AI can support post-visit engagement in tourism settings and offers practical guidance for firms seeking efficient strategies to manage online reviews and strengthen their digital service presence.

摘要

本研究探讨了生成式人工智能能否作为一种稳定且可扩展的数字互动形式，用于提升在线评论中的游客情感。通过采用准实验的双重差分法，我们分析了在英国某旅游城市五家印度餐厅收集的为期六个月的 11,393 条评论。其中一家餐厅采用 ChatGPT 生成回复，其余四家作为控制组。结果显示，由人工智能中介的回复显著改善了游客情感，且这种积极影响在社交平台（如 Google 、Facebook ）和外卖服务平台（如 Deliveroo 、 Just Eat 、UberEATS ）上表现相似。回复的特征，如语气、个性化和长度，其额外影响有限。这表明，回复的存在本身比其具体的风格特质更为重要。研究结果表明，生成式人工智能中介的回复主要通过传达组织关注度和关系合法性来影响游客情感，而非依赖回复的

ARTICLE HISTORY
Received 18 November 2025
Accepted 24 February 2026

KEYWORDS
Tourist sentiment; generative AI; Service-Dominant Logic; online reviews; quasi-experimental design

关键词
游客情感; 生成式人工智能; 服务主导逻辑; 在线评论; 准实验设计

CONTACT Waqas Ali waqas.ali@plymouth.ac.uk Plymouth Business School, University of Plymouth, Room 214, Fitzroy Building, Drake Circus, Plymouth PL4 8AA, UK

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

细微风格特征。本研究展示了人工智能如何在旅游场景中支持游客到访后的互动，并为那些寻求高效管理在线评论、强化其数字服务能力的企业提供了实践指导。

1. Introduction

The rapid integration of artificial intelligence (AI) into tourism and hospitality services has transformed not only how tourists interact with firms but also how firms manage digital feedback. Online reviews now function as highly visible narratives of the tourist experience, shaping trust, reputation, and destination appeal (Bigne et al., 2019; Filieri et al., 2015; Schuckert et al., 2015). In this environment, tourists increasingly expect service providers to engage with their reviews, influencing not only the reviewer's satisfaction but also the perceptions of prospective travellers (Mariani & Predvoditeleva, 2019; Zhang et al., 2015). Managerial responses have, therefore, been studied as mechanisms of relationship repair and brand reinforcement (Keller, 1999; Petriglieri, 2015; Umasuthan et al., 2017). Although extant research assumes that responsiveness in online review environments signals authenticity, care, and effort, it remains unclear whether such inferences persist when engagement is mediated by generative AI rather than human actors (Huang & Rust, 2021; Sparks et al., 2016).

Generative AI models such as ChatGPT can now produce contextually relevant and emotionally calibrated responses at scale (Park & Kim, 2024; Paul et al., 2023). Hospitality firms have begun adopting these tools to automate engagement with online reviews (Dwivedi et al., 2024; Sigala et al., 2024). The promise of AI lies in scalability, but its risk lies in perceived inauthenticity: tourists may view AI-generated responses as superficial automation rather than genuine relational engagement. The literature on AI in hospitality is still largely conceptual and tends to overlook the complex, dynamic, and platform-mediated realities of review ecosystems (Callarisa-Fiol et al., 2023; Mukherjee et al., 2023; Tan et al., 2025). What remains unclear is whether AI-mediated responses are accepted as credible acts of service, and whether the relational cues that matter in human responses, such as tone, personalisation, or length, retain their effectiveness when generated by AI.

To address this gap, and in line with S-D Logic, our study conceptualises generative AI not merely as a communication tool, but as a service interface whose legitimacy is inferred through observable cues such as responsiveness and contextual appropriateness in post-consumption interactions (Van Doorn et al., 2010; Vargo & Lusch, 2014b). From this perspective, AI responses may alter tourists' relational and affective evaluations, influencing not only the original reviewer but also prospective tourists who interpret these responses in public platforms. At the same time, the legitimacy of AI as a co-creating actor is uncertain. If tourists disregard AI's attempts at empathy and personalisation, this challenges existing assumptions about relational signalling in service engagement.

Our study employs a quasi-experimental Difference-in-Differences (DiD) design across five Indian restaurants in a UK tourist city located in the southwest of England. Restaurant A implemented ChatGPT-based review responses, while four matched restaurants did not. Review data from three delivery platforms (Deliveroo, Just Eat, UberEATS) and two socially oriented platforms (Google, Facebook) allow us to test whether effects are consistent

across different digital ecosystems. Beyond estimating causal effects, we examine whether traditional response cues such as tone, personalisation, and length moderate AI effectiveness. We address these using three research questions:

RQ1: Do ChatGPT-generated responses enhance tourist sentiment relative to matched controls?

RQ2: Are GAI effects consistent across delivery and social platforms, or does platform context shape outcomes?

RQ3: Do response features such as tone, personalisation, and length moderate the effectiveness of GAI responses, or are these relational cues discounted when generated by algorithms?

Our study adopts a cautious stance by recognising that perceptions of AI authorship, contextual conditions, and sentiment measurement choices can shape how AI-mediated engagement effects are understood (Mogaji et al., 2024). By addressing our research questions, this study contributes in three ways. First, it provides field-based causal evidence that generative AI can function as a co-creating actor in service ecosystems. Second, it advances theory by showing that authenticity cues central to human responses may lose salience in GAI-mediated contexts, extending debates on trust and legitimacy in S-D Logic. Third, it offers practical guidance for hospitality firms, suggesting that scalable AI engagement can enhance digital reputation but may not require complex personalisation or tone calibration to be effective.

2. Literature review

2.1 Digital customer engagement and tourist sentiment

Online review platforms have become central to how tourists evaluate hospitality services. Reviews shape destination images, influence booking decisions, and signal both functional and emotional outcomes of service encounters (Zeng & Gerritsen, 2014). Tourist sentiment, derived from review text, is a meaningful indicator of satisfaction and perceived service quality, and it influences the credibility and persuasiveness of the review for future travellers (Filieri et al., 2015; Guo et al., 2017).

Most research focuses on how tourists express sentiment, but less attention has been paid to how firms can influence sentiment through post-visit engagement. As online reviews increasingly act as interactive spaces instead of static evaluations, understanding how firms shape sentiment through visible responses becomes important for tourism operators.

2.2 Managerial responses and the rise of AI-mediated engagement

Managerial responses serve as relational signals that express attentiveness and accountability. Prior studies show that when firms respond to reviews, customers perceive higher trust, transparency, and service orientation (Kwok & Xie, 2016). For negative reviews, empathetic replies reduce reputational harm, and for positive reviews, appreciative responses strengthen relational bonds (Liu & Park, 2015; Sparks et al., 2016).

However, many tourism businesses, especially small independent operators, lack the time and staff to respond consistently (Cheng & Jin, 2019). This has motivated interest

in generative AI (GAI) as an alternative tool for scalable engagement. While early evidence suggests that AI-generated responses can match human responses in tone and clarity (Gherheș et al., 2024), other studies argue that customers may question their sincerity or authenticity (Longoni et al., 2019).

This tension raises a key question: will tourists accept GAI-mediated responses as meaningful interactions capable of shaping sentiment?

2.3 Generative AI as a service actor in value co-creation

Service research suggests that generative AI should not be viewed solely as a neutral communication tool. Rather, it is a perceptible service interface shaping customer sensemaking. Service-Dominant Logic (S-D Logic) views value as co-created through interactions between firms and customers (Vargo & Lusch, 2008, 2014a, 2014b). From this perspective, responding to online reviews even through AI can be considered a co-creating act that integrates resources, shapes perceptions, and signals relational value (Shaw et al., 2011; Sparks & Browning, 2011).

In line with SD logic, value emerges through interactions where operant resources signal responsiveness and accessibility (Vargo & Lusch, 2014a, 2014b). Customers form affective evaluations based on algorithmic engagement, particularly in routine, standardised service encounters (Castelo et al., 2019; Van Doorn et al., 2010). The AI-mediated responses can therefore function as resource-integrating mechanisms that influence tourist sentiment by signalling organisational attentiveness rather than relational depth in hospitality context (Dogru et al., 2025).

Supporters of AI in service interactions argue that customers value responsiveness more than authorship, and that AI can fulfil functional and relational roles when its communication is respectful and contextually calibrated (Belanche et al., 2021). Others challenge whether AI can authentically express empathy or personalisation, which may weaken its perceived legitimacy as a co-creating actor (Huang & Rust, 2021).

Thus, S-D Logic suggests that AI can co-create value, but authenticity concerns raise questions about the strength and boundary conditions of such effects.

2.4 Relational cues and platform context as moderators

Tone, personalisation, and response length are important relational cues in digital communication. Warm or appreciative tones are associated with higher trust, while neutral or overly formal tones may reduce perceived sincerity (Ludwig et al., 2013). Personalisation signals attentiveness and care (Blitvich et al., 2019), and longer responses often indicate greater effort and engagement (Ghasemaghaei et al., 2018).

However, when cues come from AI, their effect may differ. Some studies show that customers discount artificial empathy or scripted personalisation (Gretzel & Koo, 2021). Others find that if GAI maintains clarity and politeness, users may still respond positively (Tan et al., 2025).

Platform context also shapes expectations. Social platforms (Google, Facebook) emphasise relational communication, while delivery platforms (Deliveroo, Just Eat, Uber-EATS) prioritise transaction speed and operational efficiency (Lu et al., 2019). These

variations create different norms that may amplify or weaken the effect of responses generated through GAI.

2.5 Research gap and hypothesis development

Although GAI tools are increasingly used in tourism, few empirical studies test their real effects using real-world behavioural data. Most work remains conceptual or based on simulated scenarios (Bilgihan et al., 2024; Seyfi et al., 2025). It is still unclear whether AI responses act as credible co-creation mechanisms and whether tourists adjust sentiment in response to AI interactions. To address this gap, we draw on S-D Logic and digital engagement literature to develop the following hypotheses.

There are several reasons why AI-generated responses may increase tourist sentiment. First, S-D Logic suggests that any responsive interaction, even if automated, acts as a co-creation touchpoint that shapes how customers evaluate service experiences (Vargo & Lusch, 2008). Second, public responses signal attentiveness and accountability, which enhance trust and perceived service quality (Kwok & Xie, 2016). Third, research shows that responsiveness itself is valued, regardless of the complexity of the message (Sparks et al., 2016). Conversely, some scholars argue that customers may discount AI-generated messages due to concerns about authenticity, emotional sincerity, or perceived effort (Longoni et al., 2019). If customers interpret automated responses as mechanical or insincere, the interaction might not generate positive sentiment, limiting the overall effect. Thus, AI-generated responses can shape tourist sentiment if they are interpreted as legitimate indicators of firm engagement, signalling responsiveness and accountability rather than merely conveying information (Castelo et al., 2019; Huang & Rust, 2021). Accordingly, we hypothesise:

> H1: Restaurants that implement ChatGPT-generated responses will show a more positive increase in tourist sentiment over time compared to matched control restaurants that do not respond.

There are theoretical reasons to expect a consistent effect. First, responsiveness is a universal relational cue across digital environments, regardless of platform norms (Sparks & Browning, 2011). Second, AI responses offer clarity, politeness, and acknowledgement, which are valued across both transactional and social contexts (Belanche et al., 2021). Third, the visibility mechanism means that responses influence not only the original reviewer but also future readers across platforms. However, platform cultures differ. Social platforms emphasise emotional and narrative content, meaning AI responses might be more scrutinised for warmth and authenticity (Xie & So, 2018). Delivery platforms prioritise speed and efficiency, so longer or expressive AI responses may appear incongruent or unnecessary (Zheng & Wu, 2022). Thus, platform norms could moderate the impact. Accordingly, we hypothesise:

> H2: The positive effect of ChatGPT-generated responses on tourist sentiment will generalise across delivery-focused and socially oriented platforms.

Longer responses often signal greater effort, thoughtfulness, and engagement in computer-mediated communication (Ghasemaghaei et al., 2018). Customers may infer that the business invested more time in addressing feedback, which enhances relational value. On

the other hand, AI-generated length may be interpreted as filler or template-like content. Customers might disregard long AI responses if they perceive them as generic or artificially extended (Gretzel & Koo, 2021). This could weaken the assumed positive relationship. Accordingly, we hypothesise:

H3a: Longer AI-generated responses will be associated with higher tourist sentiment scores.

Personalisation such as addressing the reviewer by name, strengthens perceptions of attentiveness and respect, which enhances relational quality in tourism (Blitvich et al., 2019). However, personalisation may backfire when produced by AI. Customers may perceive 'artificial personalisation' as scripted or manipulative, reducing credibility (Longoni et al., 2019). If users recognise that personalisation is machine-generated, its relational value may be discounted. Thus, we hypothesise:

H3b: Personalised responses will be associated with higher tourist sentiment than generic responses.

Warm and appreciative tones have been shown to increase customer satisfaction, trust, and perceived sincerity in service communication (Ludwig et al., 2013). So, emotional cues can elevate the relational impact of responses. Yet emotional cues from AI may be perceived as artificial or inauthentic. Some studies report that users are sceptical

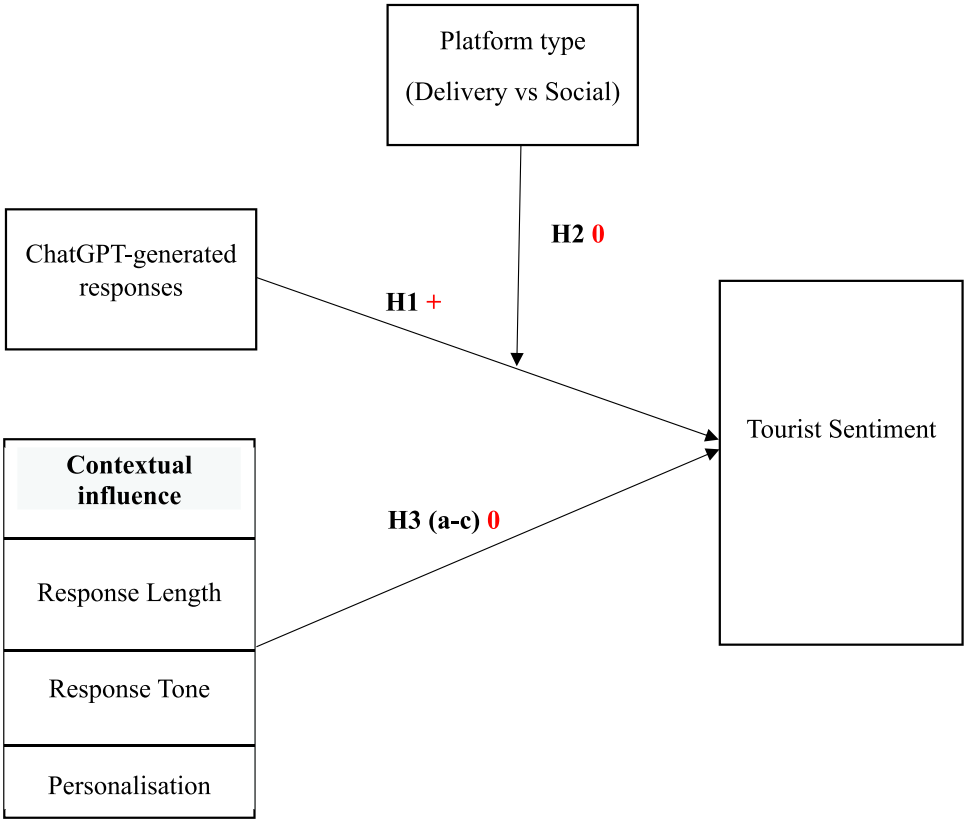

Figure 1. Theoretical framework.

of machine-generated empathy (Huang & Rust, 2021). If tourists suspect that emotional language is algorithmic, tone cues may lose effectiveness. Accordingly, we hypothesise:

> H3c: Appreciative or empathetic tones will be associated with more positive tourist sentiment than neutral or apologetic tones.

Figure 1 presents the theoretical framework for this study.

3. Methodology

3.1 Overview of methodology

This research employs a quasi-experimental framework to evaluate the impact of ChatGPT-generated responses on tourist sentiment in online restaurant reviews. The setting consists of five Indian restaurants located in a major tourist city in Devon, Southwest UK, matched on key characteristics such as menu, operating hours, size, service type, pricing, and location. Over six months, one restaurant (A) implemented the intervention by using ChatGPT to generate responses across five review platforms (Google, Facebook, Deliveroo, Just Eat, and UberEATS), while four restaurants (B–E) served as untreated controls. Data were collected from these platforms and analysed using natural language processing (NLP), Difference-in-Differences (DiD) estimation, mixed-effects regression, and complementary inferential tests, all conducted in Python. The analyses are structured into a series of interlinked components: exploratory review analysis, causal estimation of the intervention, moderation tests, robustness checks, and visualisation of longitudinal trends. Together, these components form a unified methodological approach for assessing the real-world effects of generative AI in tourism service contexts.

3.2 Data collection and analysis

Customer review data were retrospectively collected from five digital platforms named Deliveroo, Just Eat, UberEATS, Facebook, and Google Reviews for all five restaurants. The dataset covered six months, including a three-month pre-intervention phase and a three-month post-intervention phase. In total, 11,393 reviews were gathered: 3,475 before the intervention and 7,918 after. Web scraping and manual collection captured review texts, star ratings, timestamps, and owner responses. All reviews were de-identified, and data handling procedures complied with ethical standards for secondary analysis of publicly available web data. Each review was labelled by restaurant ID (A, B, C, D and E), platform (Deliveroo, Just Eat, UberEATS, Facebook, and Google Reviews), and time period (pre/post). Sentiment was measured using natural language processing (NLP) models implemented in Python, with polarity scores ranging from −1 (negative) to +1 (positive). Additional features extracted included review length, presence of keywords, and metadata such as response frequency and response delay. Analyses were conducted in Python 3.11, chosen for its reproducibility, open-source ecosystem, and compatibility with advanced NLP and statistical modelling tools (Navlani et al., 2021; Van Thieu, 2024).

Taken together, these methodological steps provide a coherent framework for evaluating the causal effect of ChatGPT-generated responses on tourist sentiment. The design integrates baseline analysis, causal estimation, moderation tests, and robustness checks,

supported by both statistical modelling and visualisation. This multi-layered approach ensures that findings are not only statistically credible but also interpretable within the real-world dynamics of online review platforms.

We conceptually assessed the robustness of the findings against alternative sentiment representations used in prior research, including emotion-based lexicons and embedding-based semantic approaches (Devlin et al., 2019; Ludwig et al., 2013). Since our analyses focus on relative changes over time between treated and control units, the estimated treatment effects are not theoretically dependent on a single sentiment model specification. This reduces concerns that the results are driven by idiosyncrasies of a particular NLP tool rather than by substantive changes in tourist sentiment.

The following sections present the stage-wise data analysis details.

3.3 Pre-intervention review analysis

The first stage of analysis characterised the baseline customer review landscape across all five contributing restaurants (A–E) during a three-month pre-intervention period. Both qualitative and quantitative approaches were used to capture review content, dominant themes (e.g. food quality, service speed, ambience), emotional tone, and sentiment polarity. Natural language processing (NLP) techniques in Python were employed, with review text tokenised using SpaCy and sentiment scored using TextBlob on a continuous scale from −1 (negative) to +1 (positive) (Alabrah et al., 2022; Tetteh & Thushara, 2023). To ensure reliability, a subset of reviews was manually coded and compared against automated sentiment classifications. This baseline analysis provided an understanding of customer expectations and discourse prior to the deployment of ChatGPT-generated responses and served as the reference point for subsequent causal estimation.

3.4 Difference-in-differences estimation of the intervention

The causal analysis applied a Difference in Differences (DiD) design to estimate the effect of ChatGPT-generated responses on tourist sentiment. Restaurant A adopted ChatGPT to respond to all reviews during the three-month intervention period, while Restaurants B to E continued their usual practices and served as controls. The restaurants were matched on operational and geographic characteristics, which created a comparable baseline and strengthened the credibility of the natural experiment (Handley et al., 2018; Shadish et al., 2002). Sentiment scores were analysed across pre and post periods, and the DiD estimator captured the differential change in sentiment for the treated restaurant relative to the average change across controls. All estimations were conducted in Python using the statsmodels package, with the treatment-by-post-interaction as the main parameter.

To further reinforce internal validity, several supplementary diagnostics were incorporated. First, formal pre-trend checks were conducted through a pre-intervention regression and an event study specification, supported by confidence intervals, to assess whether sentiment trajectories were parallel before the intervention. Second, a placebo DiD was estimated within the pre-period alone to evaluate whether spurious treatment effects appeared when no real intervention occurred. Third, a mixed effects DiD model with restaurant-level random intercepts was included to adjust for unobserved heterogeneity across units. Together, these diagnostic tests provide a more rigorous assessment of

the identifying assumptions and strengthen confidence in the causal interpretation of the intervention effects.

3.5 Platform content analysis

The next analysis examined whether the intervention effect generalised across different digital platforms. Separate DiD models were estimated for delivery-oriented platforms (Deliveroo, Just Eat, UberEATS) and socially oriented platforms (Facebook, Google) to assess whether the effectiveness of AI-generated responses varied across these ecosystems. Fixed effects for platform type were included in mixed-effects models to account for platform-specific review cultures and user engagement patterns (Giesselmann & Schmidt-Catran, 2022). To complement statistical estimation, box plots and platform-specific trend lines were generated in Python to visualise outcome variation across delivery and social platforms. This step provided insight into the robustness of the intervention effect across distinct review environments.

3.6 Relational cue (response quality) analysis

To assess whether the relational characteristics of ChatGPT-generated responses influenced outcomes, sentiment was modelled as a function of response length, presence of personalisation (e.g. use of customer name), and emotional tone. These cues were identified through regular expression analysis and vector embeddings in Python (scikit-learn, transformers). A composite measure of 'response richness' was constructed and included as a moderator in the mixed-effects DiD model. This analysis enabled a preliminary test of whether more elaborate and relationally expressive AI responses through greater length, personalised content, or empathetic tone were associated with greater improvements in tourist sentiment. The approach builds on prior research into dose–response relationships in service interaction quality (Ritz et al., 2015).

3.7 Robustness and sensitivity checks

To ensure the reliability of the estimated treatment effect, several robustness and sensitivity checks were carried out. First, placebo Difference in Differences tests were run within the pre-intervention period to verify that no artificial treatment effect appeared when no intervention had occurred, which provides additional support for the validity of the design (Eggers et al., 2024). Second, a leave-one-out analysis was performed by removing each control restaurant in turn to test whether the effect was driven by any single comparator. Third, model diagnostics, including AIC, BIC, log likelihood measures, and an inspection of residual patterns, were used to check model specification. Fourth, a winsorised DiD model was estimated to reduce the influence of extreme values in sentiment scores. Finally, temporal dependence was assessed through autocorrelation diagnostics, and a two-way fixed effects model with restaurant and week level fixed effects was estimated to adjust for unobserved heterogeneity in both units and time. Across these checks, the direction and significance of the treatment effect remained consistent, indicating that the results were stable across alternative specifications and assumptions.

3.8 Visualisation of longitudinal trends

To complement statistical analyses, several visualisations were generated to trace sentiment dynamics over time and across platforms. A spaghetti plot illustrated weekly average sentiment trajectories for all five restaurants, with a vertical marker indicating the onset of the intervention. Additional box plots compared sentiment distributions pre- and post-intervention, disaggregated by treatment group and by platform. Finally, a multi-panel summary figure integrated these outputs to depict treatment–control divergence, platform variation, and response quality comparisons (Pretus et al., 2023). These visualisations, created in Python using matplotlib and plotly, provided intuitive evidence of divergence following the AI intervention and supported the interpretation of the quantitative models.

4. Results

4.1 Pre-Intervention review characteristics results

Analysis of 3,475 pre-intervention reviews revealed that customer sentiment was generally skewed towards the positive end of the scale. As shown in Figure 2, most polarity scores fell between 0.4 and 1.0, indicating that even before the intervention, customers expressed more favourable than unfavourable experiences. This positive baseline is typical of online restaurant reviews but also establishes an important benchmark for detecting incremental improvements following the AI-driven intervention.

Lexical analysis of review texts (Figure 3) highlighted recurring themes such as food aroma, portion size, and temperature, alongside frequent references to popular Indian dishes, including chicken tikka masala and lamb rogan josh. These patterns suggest that

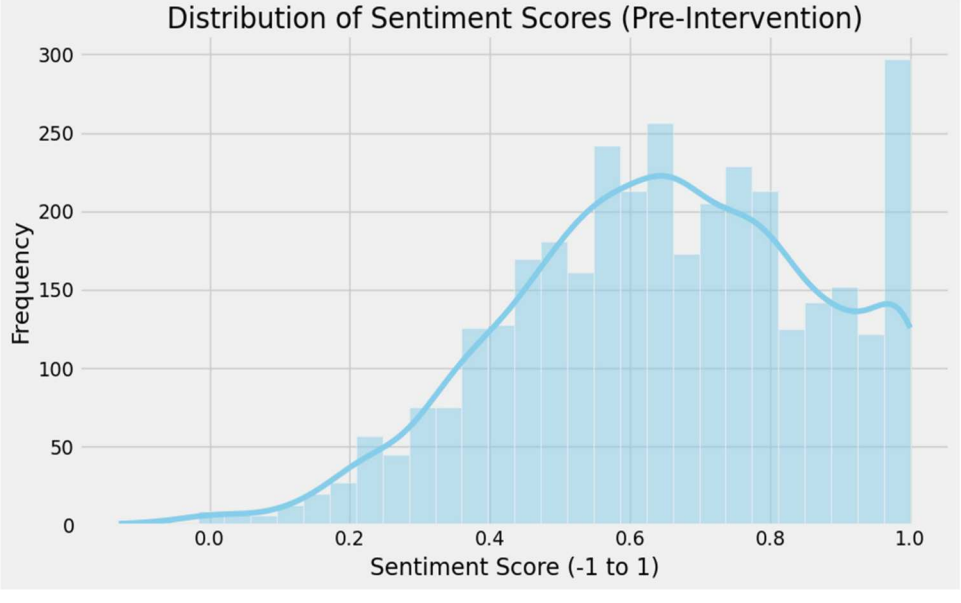

Figure 2. Sentiment score distribution.

Figure 3. Lexical analysis and key themes.

customer discourse was strongly centred on sensory qualities of the dining experience, which are known drivers of satisfaction and sentiment expression in hospitality contexts.

Descriptive statistics by restaurant (Table 1) further demonstrated that Restaurant A, the treatment group (M = 0.643, SD = 0.206), began with slightly lower average sentiment than two of the control restaurants (B: M = 0.735; E: M = 0.685). This indicates that the treated restaurant did not enjoy a pre-intervention advantage in customer sentiment, strengthening the interpretability of subsequent causal estimates.

4.2 Difference-in-differences estimation of the intervention

To evaluate the causal effect of ChatGPT-generated responses, a two-period Difference-in-Differences (DiD) model was estimated with Restaurant A as the treated unit and Restaurants B–E as controls. The dependent variable was the sentiment score of each review, measured on a continuous −1 to +1 scale. The coefficient of interest was the interaction between treatment assignment and the post-intervention period (Treatment × Post). Heteroscedasticity-robust standard errors were used to adjust for non-constant variance in residuals (Jochmans, 2022).

As shown in Table 2, the interaction term was positive and highly significant (β = 0.138, $p < .001$, 95% CI [0.118, 0.158]). This indicates that, on average, sentiment scores for

Table 1. Descriptive stats (pre-intervention sentiment by restaurant).

Restaurant	Mean Sentiment	Std Dev	Review Count (N)
A (Treated)	0.643	0.206	740
B (Control)	0.735	0.196	788
C (Control)	0.597	0.216	618
D (Control)	0.555	0.225	660
E (Control)	0.685	0.204	669

Table 2. DiD regression summary.

	Coefficient	Std_Err	z	P>|z|	CI Lower [25%]	CI Upper [95%]
Intercept	0.64790128	0.0042362	152.944583	0.00000	0.63959851	0.656204
Treatment	−0.0044148	0.0086764	−0.5088281	0.6108727	−0.0214202	0.0125906
Post	−0.002397	0.0050703	−0.4727436	0.6363961	−0.0123346	0.0075407
Treatment × Post	0.13825148	0.0102347	13.5081183	1.401E-41	0.11819184	0.1583111

Restaurant A increased by approximately 0.14 points more than those of the control restaurants after the intervention. Given the bounded nature of the sentiment scale, this represents a non-trivial improvement.

Neither the main effect of treatment assignment ($\beta = -0.004$, $p = .611$) nor the time effect ($\beta = -0.002$, $p = .636$) reached significance, which supports the assumption of parallel pre-intervention trends and rules out spurious time shocks as alternative explanations. Taken together, these results provide robust evidence that the introduction of AI-driven engagement via ChatGPT enhanced tourist sentiment in online reviews.

4.3 Parallel trends diagnostics (pre-intervention)

A formal pre-intervention trend test was conducted to confirm that the treatment and control restaurants followed similar sentiment trajectories before the introduction of AI responses. Table 3 presents the regression of sentiment on Treatment, Week_Index, and their interaction using only pre-period data (N = 3,475). The Treatment × Week_Index term is very small and statistically non-significant ($\beta = -0.0011$, $p = .642$), indicating no differential trend between treated and control restaurants prior to the intervention. The absence of a systematic pre-period divergence supports the parallel trends assumption required for a valid Difference in Differences design.

The pre-intervention event study further tests the parallel trends assumption by estimating week-specific differences between the treatment and control groups prior to AI adoption. Figure 4 reports coefficients for the 12 pre-intervention event weeks (relative to week −1). All coefficients are small, fluctuate around zero, and have confidence intervals that comfortably include zero, indicating no systematic divergence in sentiment before the intervention. This pattern reinforces that the treated and control restaurants followed comparable trajectories in the pre-period, supporting the validity of the Difference in Differences design.

4.3.1 Placebo difference in differences (pre period only)

The placebo Difference in Differences analyses further assess whether any treatment effect appears spuriously before the actual intervention. Across all platforms, and when delivery and social platforms are examined separately, the placebo interaction term

Table 3. Pre-intervention parallel trends test.

	Coefficient	Std_Err	z	P>|z|	CI Lower [25%]	CI Upper [95%]
Intercept	0.6498	0.009	71.986	0.000	0.632	0.667
Treatment	0.0034	0.019	0.184	0.854	−0.033	0.040
Week_Index	−0.0003	0.001	−0.236	0.814	−0.002	0.002
Treatment × Week_Index	−0.0011	0.002	−0.466	0.642	−0.006	0.004

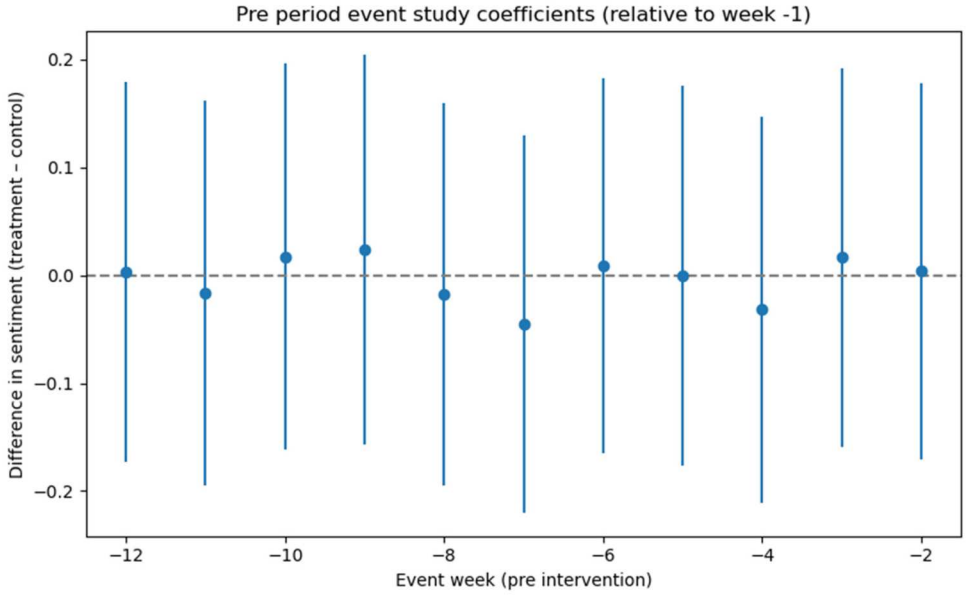

Figure 4. Pre-intervention event study coefficients.

(Treatment × Fake_Post) remains small and statistically insignificant (p values between 0.63 and 0.90). Table 4 presents all the details. This pattern indicates that no artificial 'effect' is detected when a placebo intervention date is imposed inside the pre-period. The absence of placebo effects strengthens the internal validity of the main DiD findings by confirming that the estimated post intervention increase in sentiment is unlikely to be driven by pre-existing divergence or spurious temporal patterns.

4.3.2 Mixed effects DiD with random intercepts (restaurant level)

The mixed effects Difference in Differences model, which adjusts for unobserved restaurant-level heterogeneity, confirms the robustness of the intervention effect. Details are presented in Table 5. The Treatment × Post coefficient remains positive and statistically significant ($\beta = 0.136$, $p < .001$), closely matching the magnitude found in the main DiD model. This indicates that the improvement in sentiment at the treated restaurant persists even after allowing each restaurant to have its own baseline level of sentiment. The platform type dummy shows no significant influence, suggesting that unobserved fixed characteristics at the restaurant level, rather than platform differences, account for most of the structural variation in scores. Overall, the mixed effects specification reinforces confidence in the causal interpretation by demonstrating that results are not driven by unobserved restaurant-specific factors.

Table 4. Placebo DiD estimates in the pre-period.

Model	Treatment	Fake_Post	Treatment × Fake_Post	Interpretation
All platforms	−0.0053 ($p = .628$)	0.0011 ($p = .900$)	0.0023 ($p = .900$)	No placebo effect detected
Delivery platforms	0.0067 ($p = .669$)	0.0084 ($p = .498$)	−0.0040 ($p = .872$)	No placebo effect detected
Social platforms	−0.0163 ($p = .280$)	−0.0066 ($p = .600$)	0.0084 ($p = .748$)	No placebo effect detected

Table 5. Mixed effects DiD estimates.

Term	Coefficient	Std. error	*p* Value	Interpretation
Intercept	0.32	87482.427	1	Baseline (not meaningful due to centring/model scale)
Platform_Type (Social)	0.001	0.004	0.718	No significant platform-level baseline difference
Treatment	−0.368	68436.397	1	No pre-period difference after random intercepts
Post	0	0.005	0.965	No overall post-shift unrelated to treatment
Treatment × Post	0.136	0.01	<.001	Significant treatment effect after controlling for restaurant heterogeneity
Random intercept variance	0	–	–	Minimal unobserved restaurant-level variance

4.4 Platform context analysis results

To examine whether the effect of ChatGPT-generated responses varied across review platforms, an extended DiD model was estimated with an interaction term for platform type. Delivery platforms (Deliveroo, Just Eat, UberEATS) were treated as the baseline, and social platforms (Google, Facebook) were included as a binary moderator (Schoeneberger & Rhoads, 2025).

As reported in Table 6, the main treatment effect remained positive and significant ($\beta = 0.141$, $p < .001$, 95% CI [0.118, 0.163]), confirming that ChatGPT responses improved sentiment on delivery platforms by approximately 0.14 points. The interaction term for social platforms was non-significant ($\beta = -0.004$, $p = .696$), indicating that the intervention's effectiveness did not differ across platform types.

This finding implies that the impact of AI-generated responses is platform-agnostic, extending equally to socially oriented platforms where reviews are more relational in tone and to delivery platforms where reviews are primarily transactional. The consistency of effects across digital ecosystems strengthens the external validity of the intervention and suggests that generative AI can be a scalable engagement tool for hospitality firms regardless of platform norms.

4.5 Relational cue (response quality) analysis

To examine whether the richness of ChatGPT-generated responses moderated their effectiveness, a regression model was estimated using post-intervention reviews from Restaurant A. Predictors included response length, presence of personalisation (Yes/No), and response tone (neutral, apologetic, or appreciative as reference). The dependent variable was the sentiment score of each review.

As shown in Table 7, none of the response quality indicators emerged as statistically significant predictors of sentiment. While the coefficients were directionally consistent with

Table 6. Extended DiD regression including platform-type interaction.

	Coefficient	Std_Err	z	*P*>|z|	CI Lower [25%]	CI Upper [95%]
Intercept	0.64594253	0.00476	135.586	0.0000	0.63661	0.65528
Platform_Type[T.Social]	0.004027954	0.00445	0.90434	0.36582	−0.0047	0.01276
Treatment	−0.0046061	0.00868	−0.5308	0.59558	−0.0216	0.0124
Post	−0.00248355	0.00507	−0.4897	0.62434	−0.0124	0.00746
Treatment × Post	0.140494951	0.01151	12.2061	2.9E-34	0.11794	0.16305
Treatment × Post: Platform_Type[T.Social]	−0.00404159	0.01033	−0.3911	0.6957	−0.0243	0.01621

Table 7. Dose response regression.

	Coefficient	Std_Err	z	*P*>|z|	CI Lower [25%]	CI Upper [95%]
Intercept	0.76547605	0.02881	26.5735	1E-155	0.709017398	0.821934709
Response_Personalised[T.Yes]	−0.0099076	0.00986	−1.0049	0.31493	−0.029230692	0.009415553
Response_Tone[T.Appreciative]	−0.0192018	0.01123	−1.7106	0.08716	−0.041203191	0.002799678
Response_Tone[T.Neutral]	−0.0069615	0.01146	−0.6076	0.54344	−0.029416559	0.015493624
Response_Length	0.00025346	0.00023	1.09841	0.27203	−0.000198806	0.00070573

Table 8. Illustrative examples of appreciative and neutral AI responses.

Review_Text	Response_Text	Tone
'Very generous portion of Vegetable Samosas. Will order again.'	'Thank you for your review. We are glad the portion size met your expectations.'	Appreciative
'Best Chicken Tikka Masala I have had in ages. Highly recommended.'	'Thank you for sharing your experience. It is great to know you enjoyed the dish.'	Appreciative
'The Paneer Butter Masala had an amazing aroma. Proper Indian flavour.'	'We appreciate your feedback and are pleased the flavour worked well for you.'	Appreciative
'I ordered the Chana Masala, fantastic flavours and great portion size.'	'Thank you for taking the time to share this. We are happy you liked the meal.'	Appreciative
'The Lamb Rogan Josh arrived lukewarm but still tasty.'	'Thank you for your review and for providing details about your experience.'	Neutral
'The Butter Chicken was a bit undercooked this time.'	'Thank you for your feedback. Your comments about the dish have been noted.'	Neutral
'The Palak Paneer was way too spicy for my taste.'	'Thank you for sharing your thoughts. We have taken note of your preference.'	Neutral

expectations. For example, longer responses and appreciative tones were positively associated with sentiment, their magnitudes were small, and overall explanatory power was minimal (e.g. Response_Length: $\beta = 0.0003$, $p = .272$; Response_Personalised: $\beta = -0.0099$, $p = .315$, $R^2 = 0.003$). The non-significant effects of tone are easier to understand when looking at the actual responses. As the excerpts in Table 8 show, appreciative replies typically included simple acknowledgements such as 'thank you', while neutral replies focused on acknowledging issues without adding emotional depth. Both styles remain brief and functional, which may explain why they did not produce meaningful divergence in sentiment outcomes.

These results suggest that, within this naturalistic setting, the presence of a response itself was more consequential than its stylistic qualities. In other words, tourists appeared to value the act of acknowledgement rather than the particular tone, length, or degree of personalisation. These null findings should be interpreted with caution, as the moderation analysis relies on a subset of reviews from a single treated restaurant, which limits variability and statistical power to detect small effects.

4.6 Robustness and sensitivity checks

To test the stability of the estimated treatment effect, four robustness checks were conducted: (i) Placebo test, (ii) Leave-one-out analysis, (iii) Model diagnostics, and (iv) Winsorised DiD Robustness Check.

Placebo test: A falsification test was run by introducing a 'fake' intervention date within the pre-intervention period (November). As reported in Table 9, the interaction term was statistically non-significant ($\beta = 0.0023$, $p = .900$). This indicates that no spurious treatment-like effect emerged in the absence of the actual intervention, supporting the parallel trends assumption that underpins the DiD design.

Leave-one-out analysis: To assess whether results were disproportionately influenced by a single comparator, the DiD model was re-estimated after sequentially removing each of the four control restaurants. Across all specifications (Table 10), the treatment effect remained statistically significant and similar in magnitude ($\beta \approx 0.134$–0.141, $p < .001$). This demonstrates that the findings are not driven by any individual control unit.

Model diagnostics: Residual analysis further confirmed the adequacy of the model specification. As shown in Figure 5, residuals displayed no systematic pattern or signs of

Table 9. Place DiD regression.

	Coefficient	Std_Err	z	*P*>|z|	CI Lower [25%]	CI Upper [95%]
Intercept	0.647505682	0.005316294	121.7964435	0.00000	0.637085937	0.657925426
Treatment	−0.00528106	0.010904832	−0.484286195	0.628182772	−0.026654139	0.016092019
Fake_Post	0.001109703	0.00880207	0.12607293	0.899674196	−0.016142038	0.018361444
Placebo_Interaction	0.002261343	0.018031618	0.125409876	0.90019907	−0.03307998	0.037602666

Table 10. Leave-one-out sensitivity analysis.

Excluded_Control	Treatment_Effect	*P*_Value
B	0.136942062	2.48617E-37
C	0.141194663	1.3919E-40
D	0.134456901	1.5594E-37
E	0.13905874	1.16037E-38

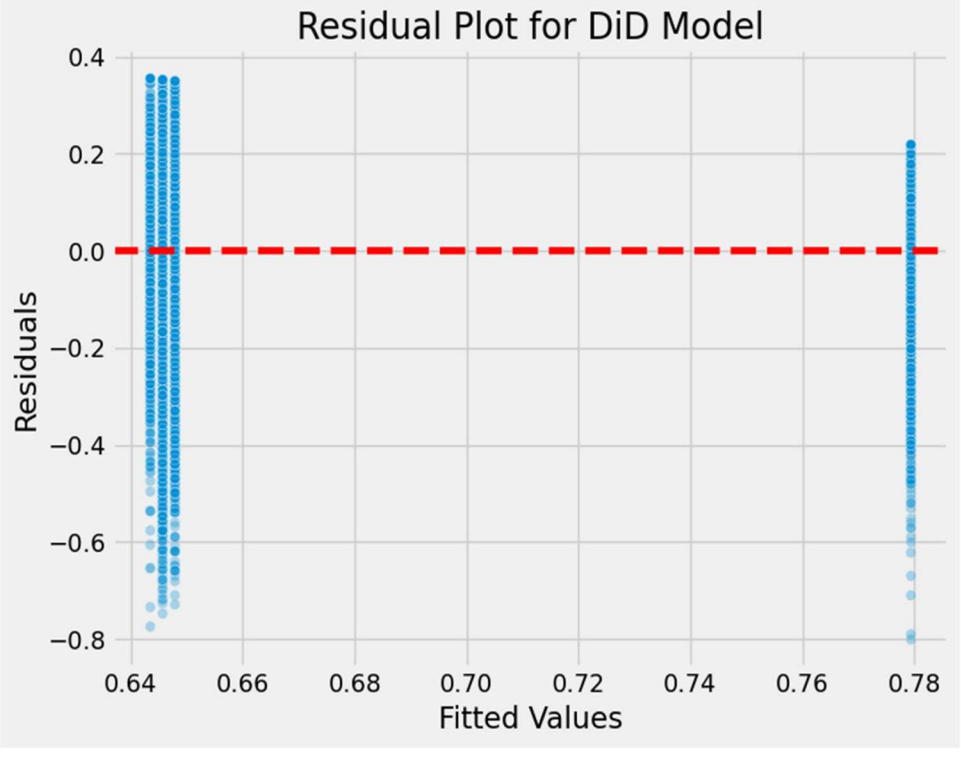

Figure 5. Residual plot for the DiD model.

heteroscedasticity, and fitted values aligned closely with expected sentiment distributions across groups. Collectively, these diagnostics confirm that the estimated treatment effect is both statistically reliable and substantively robust to common threats to internal validity (Handley et al., 2018; Shadish et al., 2002).

Winsorised DiD Robustness Check: The winsorised Difference in Differences model provides an additional robustness check by reducing the influence of extreme sentiment

Table 11. Winsorised DiD estimates.

Term	Coefficient	Std. Error	*p*-Value	Interpretation
Intercept	0.649	0.004	<.001	Baseline sentiment level
Treatment	−0.004	0.009	0.63	No pre-difference between treated and control
Post	−0.002	0.005	0.629	No general post-shift unrelated to treatment
Treatment × Post	0.137	0.01	<.001	Stable and significant treatment effect even after winsorisation

values. As reported in Table 11, after trimming the bottom and top one per cent of the sentiment distribution, the Treatment × Post coefficient remains positive and highly significant (β = 0.137, $p < .001$). This is almost identical to the main DiD estimate, indicating that the treatment effect is not driven by outliers or unusual weeks. The Treatment and Post terms remain non-significant, which further supports the parallel trends assumption because no systematic differences appear outside the true intervention period. The robustness of the effect after winsorisation strengthens confidence that the intervention consistently increased sentiment rather than reflecting noise in the data.

4.6.1 Autocorrelation diagnostics

To assess whether serial dependence in the weekly review data biased the DiD estimates, autocorrelation diagnostics were conducted. The Durbin Watson statistic is 1.995, which is close to the ideal value of 2, indicating no meaningful first-order autocorrelation in the residuals. The Breusch Godfrey LM test (four lags) also shows no evidence of residual autocorrelation (LM = 3.688, $p = 0.450$). These results suggest that the treatment effect is not driven by unmodelled temporal dependence and that the DiD specification is appropriate for the weekly structure of the data. The key interaction term remains positive and highly significant (β = 0.138, $p < .001$), further confirming the stability of the intervention effect under diagnostic checks. Table 12 has the details.

4.6.2 Two-way fixed effects estimation

To further assess the robustness of the intervention effect, a two-way fixed effects model was estimated with restaurant-specific and week-specific intercepts and clustered standard errors. This specification controls for all time-invariant restaurant characteristics and for weekly shocks common across all units. The key DiD coefficient remains positive and strongly significant (β = 0.135, $p < .001$), closely matching the main model estimate. This consistency suggests that the treatment effect is not driven by unobserved heterogeneity or dynamic differences in adoption timing. The fixed effects for individual restaurants and weeks absorb expected structural variation, while the effect of the intervention remains stable after controlling for these additional sources of bias. Pre-intervention differences between treated and control units are minimal, as shown by the small standardised mean difference (SMD = 0.021) and balanced variance ratio (0.864). Table 13 has the details.

4.7 Longitudinal visualisation of sentiment dynamics

Visual diagnostics were employed to complement the statistical analyses and trace sentiment dynamics across restaurants and platforms. The spaghetti plot of weekly average sentiment scores (Figure 6) revealed that prior to the intervention, sentiment trajectories for all restaurants were relatively stable and parallel. Following the onset of the ChatGPT

Table 12. Autocorrelation diagnostics for the DiD model.

Statistic	Value	Interpretation
Durbin Watson	1.995	No first order autocorrelation
Breusch Godfrey LM (4 lags)	3.688	$p = 0.450$, no serial dependence
Treatment × Post coefficient	0.138	Highly significant ($p < .001$), stable under diagnostics

Table 13. Two-way fixed effects DiD results.

Statistic	Value	Interpretation
DiD coefficient (Treatment × Post)	0.135	Strong and significant, consistent with main model
Restaurant FE	Yes	Controls for time-invariant group differences
Week FE	Yes	Controls for common weekly shocks
Standard errors	Clustered	Adjusted for within-group correlation
Pre-period SMD	0.021	Very small imbalance between groups
Variance ratio	0.864	Balanced dispersion across groups

intervention in December 2024, Restaurant A (treatment) exhibited a marked and sustained upward trend, while control restaurants (B–E) continued on flat or declining paths. By January 2025, Restaurant A had surpassed all control units in weekly sentiment levels, providing strong visual evidence of a treatment effect.

The box plots in Figure 7 further illustrate this divergence. For Restaurant A (Treatment), sentiment distributions shifted notably rightward between pre- and post-intervention periods, with increases in both median and interquartile range. Control restaurants (B–E), in contrast, displayed negligible changes, reinforcing the Difference-in-Differences (DiD) findings reported before.

When disaggregated by platform type (Figure 8), both socially oriented platforms (Google, Facebook) and delivery-oriented platforms (Deliveroo, Just Eat, UberEATS) showed comparable improvements in sentiment post-intervention. The consistency of these directional gains affirms the platform-agnostic nature of the treatment effect, in line with the moderation analysis results presented above in section 4.4.

Figure 9 provides a multi-panel summary of the intervention effects, integrating treatment–control comparisons, effect sizes, and moderation analyses.

Panel A compares post-intervention sentiment distributions between the treatment restaurant (A) and the pooled control group (B–E). The violin plots show that Restaurant A's distribution shifted upward, with a higher median and a greater density of positive scores. In contrast, the control group remained centred around its pre-intervention

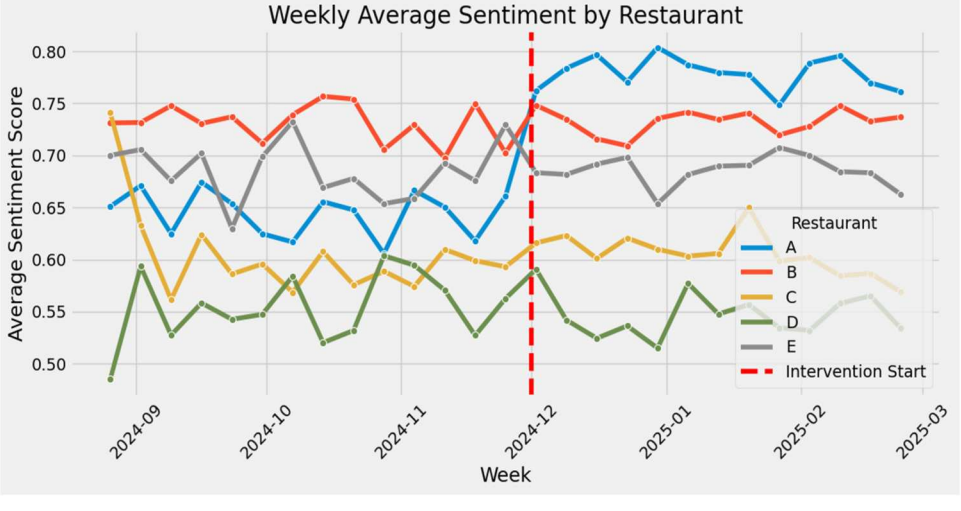

Figure 6. Weekly average sentiment by restaurant.

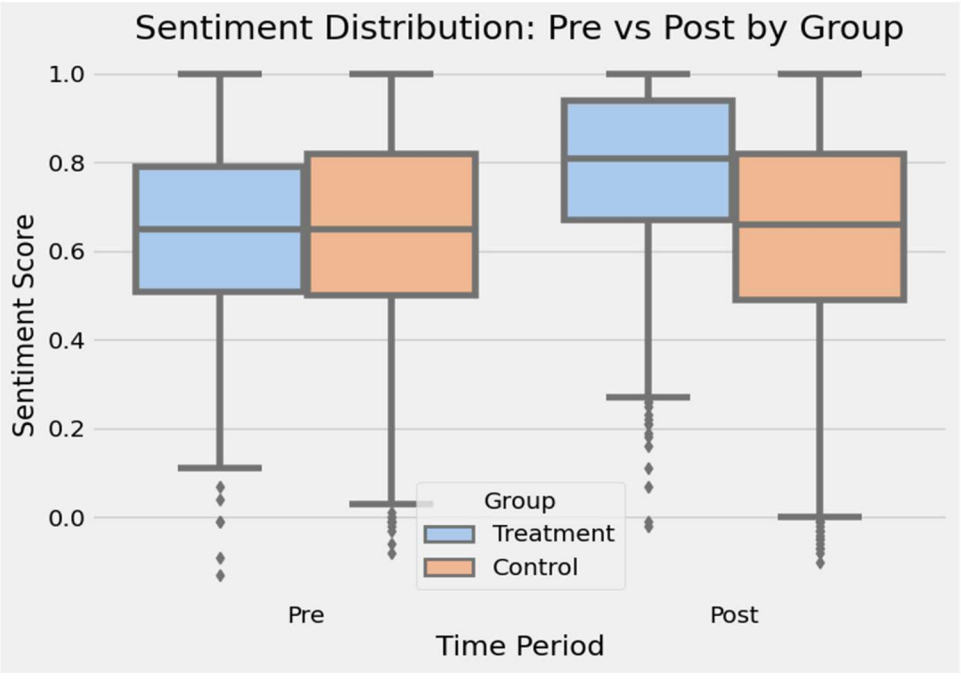

Figure 7. Sentiment distribution: pre vs post by the group.

baseline. This visualisation reinforces the DiD finding that the treatment restaurant experienced a unique sentiment gain following the intervention.

Panel B presents pre–post effect sizes (Cohen's d) for each restaurant. Restaurant A exhibited a large effect size of approximately 0.7, consistent with a substantial treatment

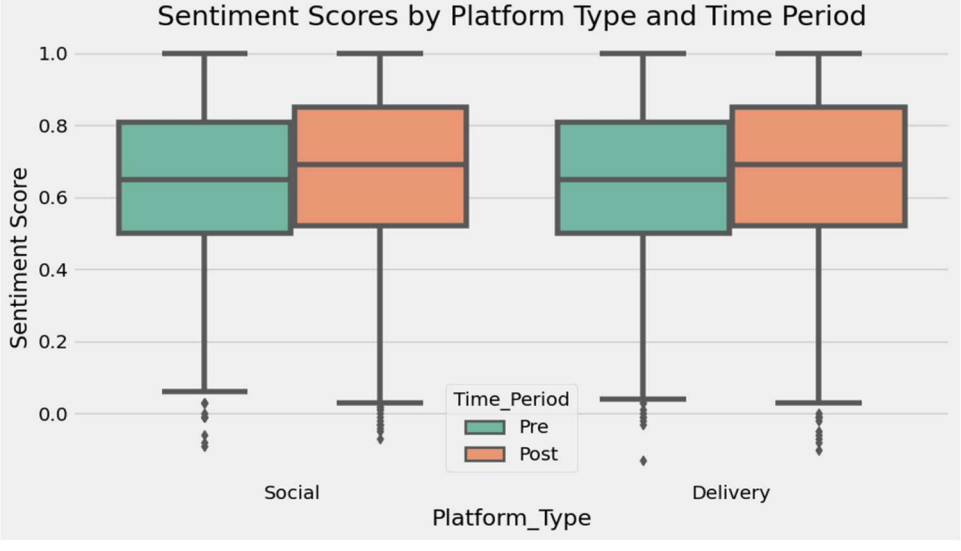

Figure 8. Sentiment scores by platform type and time period.

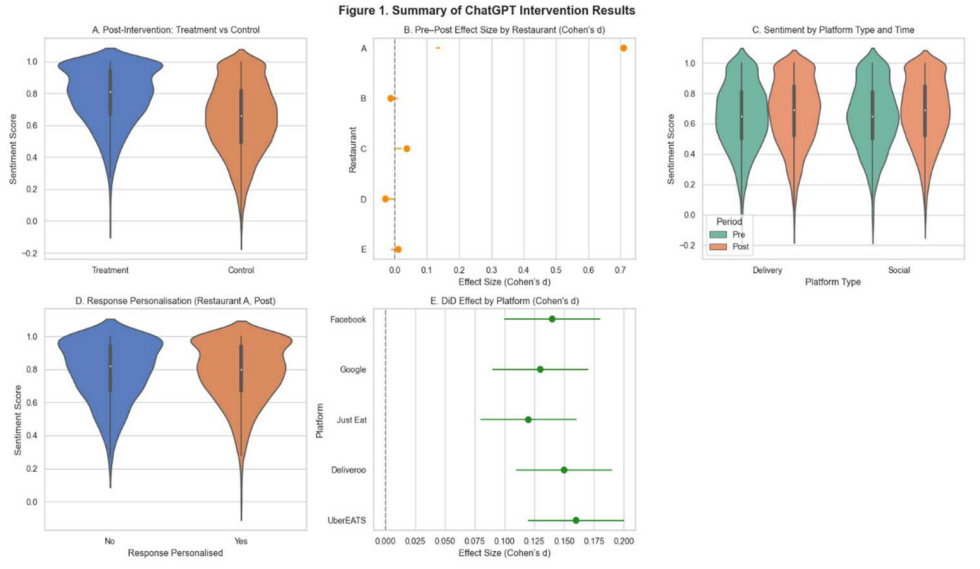

Figure 9. Multi-panel summary figure.

effect. By contrast, all four control restaurants displayed negligible effect sizes near zero, ruling out the possibility that broader contextual changes (e.g. seasonality or market shifts) were driving sentiment improvements.

Panel C disaggregates sentiment by platform type and time period. Both delivery platforms (Deliveroo, Just Eat, UberEATS) and social platforms (Google, Facebook) showed clear upward shifts in the post-intervention period, confirming that the observed treatment effect generalised across distinct review ecosystems. This finding complements the regression analysis reported in section 4.2, which showed that platform type did not significantly moderate the effect.

Panel D examines the role of personalisation within Restaurant A's responses during the intervention period. Sentiment scores were broadly similar regardless of whether responses were personalised (e.g. using the reviewer's name) or generic, suggesting that the act of responding itself was more influential than the specific framing of the reply. This aligns with section 4.3, where regression models indicated limited moderating effects of response style.

Panel E reports Difference-in-Differences effect sizes by platform. Estimates ranged from 0.12 to 0.16 across all five platforms, with overlapping confidence intervals. This consistency indicates that the intervention produced a robust and comparable uplift in sentiment across both delivery- and social-oriented environments.

Taken together, the five panels demonstrate that the positive impact of ChatGPT-generated responses was large in magnitude (Panel B), clearly visible in distributional shifts (Panel A), consistent across platforms (Panels C and E), and relatively insensitive to stylistic response variation (Panel D). These findings build a strong case for the robustness and generalisability of the intervention.

5. Discussion

The results of this research provide field-based evidence that ChatGPT-generated responses can meaningfully influence tourist sentiment in online reviews. Using a Difference in Differences approach, the study shows that Restaurant A experienced a stronger increase in sentiment after adopting AI responses compared to the matched control restaurants. This offers support for the argument that generative AI can operate as a credible engagement tool and that tourists react positively when firms acknowledge their feedback in public digital environments.

H1 examined whether AI-generated responses would increase sentiment more for the treated restaurant compared to the controls. The DiD analysis showed a statistically significant treatment effect of around 0.14 points ($p < .001$), indicating that the intervention produced a measurable uplift in sentiment. This finding supports H1 and aligns with work showing that managerial responses enhance trust, perceived quality, and booking intention in hospitality settings (Kwok & Xie, 2016). From a Service Dominant Logic perspective, the result confirms that value can be co-created not only during the on-site encounter but also through post-consumption dialogue on review platforms (Shaw et al., 2011; Vargo & Lusch, 2008). At the same time, the effect is moderate rather than dramatic, suggesting an incremental uplift in sentiment rather than a complete transformation of customer evaluations. This nuance is consistent with studies that find digital interventions often adjust perceptions at the margin rather than overturning underlying service experiences (Sparks & Browning, 2011).

H2 tested whether the effect of AI responses would generalise across delivery-focused and socially oriented platforms. The interaction between the treatment effect and platform type was not statistically significant ($p = .696$), so H2 is not supported in a strict statistical sense. However, the estimated effects were similar in size across delivery and social platforms, indicating that the pattern of improvement does not differ strongly between ecosystems. This is interesting in light of research that highlights the distinct review cultures on social versus transactional platforms (Lu et al., 2019; Xie & So, 2018). On social sites, users often expect richer narrative engagement, whereas on delivery platforms, they emphasise speed and operational performance (Zheng & Wu, 2022). The present findings suggest that, despite these differences, the basic function of visible acknowledgement retains value in both contexts. Tourists may treat responsiveness as a general norm of digital hospitality, even if their expectations for content depth differ. At the same time, because the interaction term is non-significant, these results should be interpreted as indicative of similar patterns rather than as proof that platform norms never matter. More focused designs are needed to test how far platform culture can constrain or amplify AI-mediated engagement.

H3a–c focused on whether relational cues in the AI responses, length, personalisation, and tone moderated the sentiment outcome. None of these variables were statistically significant predictors when tested within the treated restaurant during the intervention period. While this contrasts with prior findings from human-authored responses (Blitvich et al., 2019; Ghasemaghaei et al., 2018; Ludwig et al., 2013), these null effects should be interpreted with caution, as the moderation analysis is based on a single treated unit and may be underpowered to detect small stylistic effects. A conservative interpretation is that, in this setting, tourists responded primarily to the presence of a reply rather than

to fine-grained differences in wording. This interpretation is consistent with research on artificial empathy, where emotionally expressive AI messages are sometimes viewed as formulaic or less authentic than human ones (Gretzel & Koo, 2021; Huang & Rust, 2021). The weak and non-robust tendency for appreciative tone to perform slightly better suggests that emotional cues are not irrelevant, but that their impact may be constrained when the author is an AI tool.

Taken together, the pattern of findings suggests that ChatGPT can act as a scalable co-creating actor that enhances sentiment through consistent and visible engagement, rather than through fine-grained tailoring of message style. The supported effect for H1 shows that AI responses can extend the co-creation process into the post-visit phase, in line with S D Logic (Shaw et al., 2011; Vargo & Lusch, 2008). The non-significant moderation for H2 indicates that this uplift appears across both social and delivery platforms, although future work should examine platform conditions more explicitly. The null results for H3a–c challenge a common assumption in the engagement literature that 'more personalisation and richer tone are always better' (Blitvich et al., 2019; Ghasemaghaei et al., 2018). Instead, the findings point to a boundary condition in AI-mediated settings: tourists seem to reward responsiveness itself, while being more cautious about emotional or personalised cues that may be perceived as machine-generated. Overall, the results portray AI not as a full substitute for human depth, but as a reliable, scalable contributor to post-consumption value creation in hospitality reviews.

While our findings demonstrate that ChatGPT-generated responses are associated with improved tourist sentiment, they also raise important ethical and perceptual questions regarding transparency and trust. One key potential issue is if tourists later discover that managerial responses were generated by AI instead of human staff. Prior research suggests that perceptions of authenticity and sincerity play a critical role in shaping trust in service encounters (Gretzel & Koo, 2021; Kim & Kim, 2020). If AI's role is revealed retrospectively, tourists may reconsider earlier interactions as less genuine, which could reduce trust and undermine relational value.

This concern aligns with Mogaji et al. (2024), who argue that traditional technology acceptance frameworks such as the Technology Acceptance Model (TAM) are increasingly insufficient for explaining user responses to generative AI. Unlike earlier information systems, generative AI introduces ambiguity around agency, authorship, and intent. Acceptance is therefore not solely determined by perceived usefulness or ease of use, but also by ethical judgements and transparency. Therefore, acceptance becomes dynamic and context-dependent rather than stable.

6. Conclusion, implications, and research directions

6.1 Conclusion

This research examined whether ChatGPT-generated responses can shape tourist sentiment in online restaurant reviews across five major digital platforms. Using a field-based quasi-experimental Difference in Differences design, the study provides evidence that the restaurant adopting AI-generated responses experienced a clearer improvement in sentiment compared with similar control restaurants. The effect was consistent across delivery-focused and socially oriented platforms, suggesting that visible responsiveness

matters across different digital environments. At the same time, the study found that specific features of the responses, such as tone, length, and personalisation did not significantly change sentiment outcomes. Overall, the findings show that generative AI can support routine digital engagement by helping firms acknowledge customers in a timely and consistent way. The results should be interpreted as incremental and context-specific rather than transformative. The study offers a practical demonstration of how AI-supported responses may influence sentiment, but it does not claim to redefine broader theories of value creation.

The findings of this study should be interpreted in light of the study's empirical scope, by focusing on the five comparable restaurants within a single tourist city, the research prioritises internal validity and causal clarity over broad representativeness. In this sense, the study is best understood as a proof of concept demonstrating that generative AI-mediated review responses can influence tourist sentiment in real world digital environment. Rather than establishing a generalised pattern across the hospitality sector, the results provide a foundation benchmark that invites replication and extension in larger and more diverse tourism contexts.

6.2 Theoretical implications

This study offers several theoretical implications that advance current understanding of AI-mediated interactions in hospitality and tourism. Our research clarifies the boundary conditions under which generative AI operates as a co-creating actor in service ecosystems. The findings suggest that tourists primarily perceive GAI-generated responses as an extension of the restaurant's service interface rather than as an autonomous relational agent. Aligned with prior work, algorithmic agency is accepted when tasks are routine and evaluative, but not when emotional authenticity is central (Castelo et al., 2019; Dietvorst et al., 2015). Theoretically, this positions generative AI as an operant resource that supports value co-creation through signalling responsiveness, thereby refining S-D logic by specifying when AI is recognised as a legitimate contributor to relational value (Lin et al., 2025).

Algorithmic authenticity in AI-mediated service interactions can be understood as a perceptual inference rather than an intrinsic property of the technology. Tourists assess the sincerity, effort, and legitimacy of AI-generated responses based on observable cues such as responsiveness and contextual appropriateness, even in the absence of human authorship (Foroughi et al., 2026; Huang & Rust, 2021). Such perceptions shape acceptance by signalling whether algorithmic engagement aligns with expected service norms, thereby influencing affective outcomes such as sentiment (Castelo et al., 2019; Van Doorn et al., 2010).

The findings extend existing work on digital service engagement by showing that AI-generated responses can operate as effective interactional touchpoints that shape post visit sentiment. While prior research has largely assumed that relational value online must be human-driven (Sparks & Browning, 2011; Umasuthan et al., 2017), the present study demonstrates that tourists still respond positively to acknowledgement even when the responder is an algorithm (Foroughi et al., 2024). This contributes to a growing stream of research suggesting that AI can take on partial relational roles within service encounters (Ameen et al., 2023; Belanche et al., 2021; Huang & Rust, 2021).

The study provides empirical support for applying S-D logic in AI-enabled environments by illustrating how value can be co-shaped through visible digital interactions after consumption. Rather than arguing for a revision of S-D logic, the study clarifies how AI operates as a mediating resource within existing value co-creation processes. The results show that tourists interpret AI responses as indicators of effort, accessibility, and attentiveness, all elements that contribute to relational value in service ecosystems (Shaw et al., 2011; Vargo & Lusch, 2008). This complements recent theoretical discussions on the expanding role of non-human actors in co-creation (Ostern et al., 2024).

The absence of moderating effects for tone, personalisation, and length reveals an important theoretical nuance. Prior work suggests that relational cues shape emotional evaluations in digital communication (Ghasemaghaei et al., 2018). Our results show that such cues may carry less weight when customers suspect automation or when responses are perceived as formulaic. This contributes to emerging debates about the limits of artificial empathy and how authenticity perceptions influence the effectiveness of AI-mediated interactions (Gretzel & Koo, 2021). The findings, therefore, identify a boundary condition within AI-supported engagement: presence and consistency may be more influential than stylistic richness.

Our research captures tourist sentiment in a naturalistic setting where GAI authorship was not explicitly disclosed. The positive sentiment effects observed here may therefore reflect an implicit acceptance of AI-mediated engagement, driven by responsiveness and visibility rather than by informed consent. Should disclosure become mandatory or more common, tourists' evaluations may shift, particularly if they feel misled. These dynamics suggest that traditional acceptance models may only partially explain tourist responses to generative AI.

Finally, TAM-based approaches help account for functional acceptance; they do not fully capture post-discovery reactions, moral evaluations, or legitimacy judgments. Rather than signalling the end of acceptance models, as suggested by Mogaji et al. (2024), our findings support the need for their extension to incorporate ethical transparency, temporal perception, and authenticity cues. From an S-D logic perspective, this raises a critical boundary condition: AI can function as a co-creating actor only insofar as its role is perceived as legitimate within the service ecosystem. If tourists perceive AI use as deceptive or inauthentic, value co-creation may be weakened.

6.3 Practical implications

The findings offer several practical insights for tourism operators and platform managers. First, the results indicate that generative AI can help small- and medium-sized hospitality businesses maintain a visible and consistent presence across multiple review platforms. This may be particularly useful for firms with limited staff who struggle to respond to high review volumes.

Second, the study shows that the act of responding appears more influential than specific stylistic features. This means firms do not need to craft lengthy or highly personalised messages to improve sentiment; brief and polite acknowledgements generated through AI can be sufficient.

Third, the results suggest that AI tools can support multi-platform engagement, as the sentiment gains were consistent across both delivery and social channels. Firms operating

in fragmented digital environments may therefore use AI to streamline response management.

Fourth, although AI responses are effective for routine interactions, sensitive or complaint-related reviews may still require human judgment. Managers should therefore adopt a hybrid approach, allowing AI to handle routine tasks while reserving human input for complex or emotionally charged cases.

Fifth, our results imply that managers should carefully consider disclosure strategies when using generative AI, as short-term gains of efficiency may be offset by longer-term reduction of trust if tourists perceive AI use as deceptive rather than supportive.

Sixth, destination managers and platform developers may consider integrating AI-assisted response features into their systems, providing operators with accessible tools to enhance visitor engagement and maintain transparency in digital reputation management.

Finally, practitioners should interpret these implications with an awareness of contextual boundaries. While the findings suggest that GAI-generated review responses can enhance sentiment through visible acknowledgement, their effectiveness in other hospitality settings – such as hotels, attractions, or destination-level platforms – may depend on differences in service complexity, customer expectations, and brand positioning. As such, the results should be seen as indicative rather than definitive, providing guidance for experimentation rather than a universal engagement prescription.

6.4 Limitations and future research

Several limitations define the scope of this study and suggest directions for future research. First, the sample is limited to five Indian restaurants in a single UK city. While these constraints external validity, the narrow setting reduces contextual heterogeneity and strengthens internal validity, allowing the study to function as a proof of concept demonstrating whether GAI-generated review responses can influence tourist sentiment. Future research should replicate this design across larger samples, other hospitality sectors (e.g. hotels or attractions), and multi-country contexts. In addition, the moderation analysis is based on responses from a single treated unit, which may limit statistical power and the ability to detect small but meaningful stylistic effects. Second, the study cannot determine whether tourists recognised the responses as AI-generated, which may shape how they interpret sincerity or effort. Future experiments or surveys could explore how disclosure influences perceived authenticity. Third, sentiment scores were generated through NLP models that capture general polarity but may miss subtle emotional cues. Mixed methods approaches could provide richer insight. Fourth, the study focuses on short term post intervention effects. Longitudinal work could test whether the influence of AI responses persists, declines, or changes as users become more familiar with automated communication. Fourth, future research should explore service situations where emotional intelligence is more central, such as complaints or crisis scenarios, to understand when human input remains essential. Fifth, theoretically and practically, it is important to examine how explicit disclosure of GAI-generated responses influences tourist trust, sentiment, and perceived authenticity, particularly in comparison with undisclosed or hybrid human–AI engagement strategies (Prakash et al., 2023). Sixth, future studies should investigate the underlying mechanisms that link AI generated responses

with tourist sentiment. Constructs such as perceived attentiveness, relational legitimacy, or authenticity cannot be measured using naturally occurring reviews and require primary data through surveys or experiments. Examining these pathways would provide deeper insight into how and why tourists interpret AI mediated communication in digital service contexts. Seventh, future studies should explore domain specific sentiment modelling. The present analysis relied on a general-purpose sentiment tool (TextBlob), which cannot fully capture cultural cues, language nuance, or context specific expressions in tourist reviews. Developing or fine tuning a sentiment model trained on hospitality or tourism review corpora, supported by manual annotation, would help improve precision and reduce semantic bias. This represents an important methodological extension for future research. Finally, although sentiment was operationalised using polarity scores, future research could extend this approach. For instance, future researchers should incorporate transformer-based emotion models or multidimensional affective representations to capture emotional intensity and nuance more fully (Devlin et al., 2019; Ludwig et al., 2013).

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Waqas Ali http://orcid.org/0000-0002-6200-8580
Dulekha Kasturiratne http://orcid.org/0009-0000-1959-0918
Irfan Ameer http://orcid.org/0000-0002-2379-7863
Shikhar Bhaskar http://orcid.org/0000-0002-8835-1918

References

Alabrah, A., Alawadh, H. M., Okon, O. D., Meraj, T., & Rauf, H. T. (2022). Gulf countries' citizens' acceptance of COVID-19 vaccines—A machine learning approach. *Mathematics*, *10*(3), 467. https://doi.org/10.3390/math10030467

Ameen, N., Viglia, G., & Altinay, L. (2023). Revolutionizing services with cutting-edge technologies post major exogenous shocks. *The Service Industries Journal*, *43*(3-4), 125–133. https://doi.org/10.1080/02642069.2023.2185934

Belanche, D., Casaló, L. V., Flavián, M., & Ibáñez-Sánchez, S. (2021). Understanding influencer marketing: The role of congruence between influencers, products and consumers. *Journal of Business Research*, *132*, 186–195. https://doi.org/10.1016/j.jbusres.2021.03.067

Bigne, E., Ruiz, C., & Curras-Perez, R. (2019). Destination appeal through digitalized comments. *Journal of Business Research*, *101*, 447–453. https://doi.org/10.1016/j.jbusres.2019.01.020

Bilgihan, A., Dogru, T., Hanks, L., Line, N., & Mody, M. (2024). The GAI marketing model: A conceptual framework and future research directions. *International Journal of Hospitality Management*, *123*, 103929. https://doi.org/10.1016/j.ijhm.2024.103929

Blitvich, P. G. C., Fernández-Amaya, L., & de la O Hernández-López, M. (2019). An introduction to technology mediated service encounters. In P. G. C. Blitvich, L. Fernández Amaya, and M. de la O Hernández López (Eds), *Technology mediated service encounters* (pp. 1–14). John Benjamins Publishing Company.

Callarisa-Fiol, L. J., Moliner-Tena, M. A., Rodriguez-Artola, R., & Sanchez-Garcia, J. (2023). Entrepreneurship innovation using social robots in tourism: A social listening study. *Review of Managerial Science*, *17*(8), 2945–2971. https://doi.org/10.1007/s11846-023-00646-9

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research*, *56*(5), 809–825. https://doi.org/10.1177/0022243719851788

Cheng, M., & Jin, X. (2019). What do Airbnb users care about? An analysis of online review comments. *International Journal of Hospitality Management*, *76*, 58–70. https://doi.org/10.1016/j.ijhm.2018.04.004

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT* (pp. 4171–4186). https://doi.org/10.18653/v1/N19-1423

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126. https://doi.org/10.1037/xge0000033

Dogru, T., Line, N., Mody, M., Hanks, L., Abbott, J., Acikgoz, F., Assaf, A., Bakir, S., Berbekova, A., Bilgihan, A., Dalton, A., Erkmen, E., Geronasso, M., Gomez, D., Graves, S., Iskender, A., Ivanov, S., Kizildag, M., Lee, M., … Zhang, T. (2025). Generative artificial intelligence in the hospitality and tourism industry: Developing a framework for future research. *Journal of Hospitality & Tourism Research*, *49*(2), 235–253. https://doi.org/10.1177/10963480231188663

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

Eggers, A. C., Tuñón, G., & Dafoe, A. (2024). Placebo tests for causal inference. *American Journal of Political Science*, *68*(3), 1106–1121. https://doi.org/10.1111/ajps.12818

Filieri, R., Alguezaui, S., & McLeay, F. (2015). Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. *Tourism Management*, *51*, 174–185. https://doi.org/10.1016/j.tourman.2015.05.007

Foroughi, B., Ghobakhloo, M., Wen, J., & Fathi, M. (2026). Sustained use of generative AI for shopping: A PLS-ANN analysis. *The Service Industries Journal*, *46*(3-4), 250–283. https://doi.org/10.1080/02642069.2025.2485983

Foroughi, B., Huy, T. Q., Iranmanesh, M., Ghobakhloo, M., Rejeb, A., & Nikbin, D. (2024). Why users-continue E-commerce chatbots? Insights from PLS-fsQCA-NCA approach. *The Service Industries Journal*, *45*(11-12), 1–31. https://doi.org/10.1080/02642069.2024.2371910

Ghasemaghaei, M., Eslami, S. P., Deal, K., & Hassanein, K. (2018). Reviews' length and sentiment as correlates of online reviews' ratings. *Internet Research*, *28*(3), 544–563. https://doi.org/10.1108/IntR-12-2016-0394

Gherheș, V., Fărcașiu, M. A., & Cernicova-Buca, M. (2024). Are ChatGPT-generated headlines better attention grabbers than human-authored ones? An assessment of salient features driving engagement with online media. *Journalism and Media*, *5*(4), 1817–1835. https://doi.org/10.3390/journalmedia5040110

Giesselmann, M., & Schmidt-Catran, A. W. (2022). Interactions in fixed effects regression models. *Sociological Methods & Research*, *51*(3), 1100–1127. https://doi.org/10.1177/0049124120914934

Gretzel, U., & Koo, C. (2021). Smart tourism cities: A duality of place where technology supports the convergence of touristic and residential experiences. *Asia Pacific Journal of Tourism Research*, *26*(4), 352–364. https://doi.org/10.1080/10941665.2021.1897636

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent Dirichlet allocation. *Tourism Management*, *59*, 467–483. https://doi.org/10.1016/j.tourman.2016.09.009

Handley, M. A., Lyles, C. R., McCulloch, C., & Cattamanchi, A. (2018). Selecting and improving quasi-experimental designs in effectiveness and implementation research. *Annual Review of Public Health*, *39*(1), 5–25. https://doi.org/10.1146/annurev-publhealth-040617-014128

Huang, M. H., & Rust, R. T. (2021). Engaged to a robot? The role of AI in service. *Journal of Service Research*, *24*(1), 30–41. https://doi.org/10.1177/1094670520902266

Jochmans, K. (2022). Heteroscedasticity-robust inference in linear regression models with many covariates. *Journal of the American Statistical Association*, *117*(538), 887–896. https://doi.org/10.1080/01621459.2020.1831924

Keller, K. L. (1999). Managing brands for the long run: Brand reinforcement and revitalization strategies. *California Management Review*, *41*(3), 102–124. https://doi.org/10.2307/41165999

Kim, M., & Kim, J. (2020). The influence of authenticity of online reviews on trust formation among travelers. *Journal of Travel Research*, *59*(5), 763–776. https://doi.org/10.1177/0047287519868307

Kwok, L., & Xie, K. L. (2016). Factors contributing to the helpfulness of online hotel reviews: Does manager response play a role? *International Journal of Contemporary Hospitality Management*, *28*(10), 2156–2177. https://doi.org/10.1108/IJCHM-03-2015-0107

Lin, K. J., Zhang, K., Chen, S., Ye, H., Li, X., & Law, R. (2025). Value formation in AI-integrated service system: Review and implications on hospitality and tourism research. *The Service Industries Journal*, *45*(7-8), 635–656. https://doi.org/10.1080/02642069.2024.2346235

Liu, Z., & Park, S. (2015). What makes a useful online review? Implication for travel product websites. *Tourism Management*, *47*, 140–151. https://doi.org/10.1016/j.tourman.2014.09.020

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research*, *46*(4), 629–650. https://doi.org/10.1093/jcr/ucz013

Lu, L., Cai, R., & Gursoy, D. (2019). Developing and validating a service robot integration willingness scale. *International Journal of Hospitality Management*, *80*, 36–51. https://doi.org/10.1016/j.ijhm.2019.01.005

Ludwig, S., De Ruyter, K., Friedman, M., Brüggen, E. C., Wetzels, M., & Pfann, G. (2013). More than words: The influence of affective content and linguistic style matches in online reviews on conversion rates. *Journal of Marketing*, *77*(1), 87–103. https://doi.org/10.1509/jm.11.0560

Mariani, M., & Predvoditeleva, M. (2019). How do online reviewers' cultural traits and perceived experience influence hotel online ratings? An empirical analysis of the muscovite hotel sector. *International Journal of Contemporary Hospitality Management*, *31*(12), 4543–4573. https://doi.org/10.1108/IJCHM-11-2018-0927

Mogaji, E., Viglia, G., Srivastava, P., & Dwivedi, Y. K. (2024). Is it the end of the technology acceptance model in the era of generative artificial intelligence? *International Journal of Contemporary Hospitality Management*, *36*(10), 3324–3339. https://doi.org/10.1108/IJCHM-08-2023-1271

Mukherjee, S., Baral, M. M., Nagariya, R., Chittipaka, V., Rao, U. V. A., & Rao, K. S. (2023). Systematic literature review and future research directions for service robots in hospitality and tourism industries. *The Service Industries Journal*, *43*(15-16), 1083–1116. https://doi.org/10.1080/02642069.2023.2243592

Navlani, A., Fandango, A., & Idris, I. (2021). *Python data analysis: Perform data collection, data processing, wrangling, visualization, and model building using python*. Packt Publishing Ltd.

Ostern, N. K., Knickrehm, C., Voss, M., & Kelch, Y. (2024). Human-GAI co-creation: Using a process theory lens to understand human-GAI interaction in creative problem-solving.

Park, C., & Kim, J. (2024). Exploring affective representations in emotional narratives: An exploratory study comparing ChatGPT and human responses. *Cyberpsychology, Behavior, and Social Networking*, *27*(10), 736–741. https://doi.org/10.1089/cyber.2024.0100

Paul, J., Ueno, A., & Dennis, C. (2023). ChatGPT and consumers: Benefits, pitfalls and future research agenda. *International Journal of Consumer Studies*, *47*(4), 1213–1225. https://doi.org/10.1111/ijcs.12928

Petriglieri, J. L. (2015). Co-creating relationship repair. *Administrative Science Quarterly*, *60*(3), 518–557. https://doi.org/10.1177/0001839215579234

Prakash, A. V., Joshi, A., Nim, S., & Das, S. (2023). Determinants and consequences of trust in AI-based customer service chatbots. *The Service Industries Journal*, *43*(9-10), 642–675. https://doi.org/10.1080/02642069.2023.2166493

Pretus, C., Servin-Barthet, C., Harris, E. A., Brady, W. J., Vilarroya, O., & Van Bavel, J. J. (2023). The role of political devotion in sharing partisan misinformation and resistance to fact-checking. *Journal of Experimental Psychology: General*, *152*(11), 3116–3134. https://doi.org/10.1037/xge0001436

Ritz, C., Baty, F., Streibig, J. C., & Gerhard, D. (2015). Dose-response analysis using R. *PLoS One*, *10*(12), e0146021. https://doi.org/10.1371/journal.pone.0146021

Schoeneberger, J. A., & Rhoads, C. (2025). Analysis of regression discontinuity designs with a binary moderating variable. *American Journal of Evaluation*, *46*(1), 145–165. https://doi.org/10.1177/10982140241260384

Schuckert, M., Liu, X., & Law, R. (2015). Hospitality and tourism online reviews: Recent trends and future directions. *Journal of Travel & Tourism Marketing*, *32*(5), 608–621. https://doi.org/10.1080/10548408.2014.933154

Seyfi, S., Kim, M. J., Nazifi, A., Murdy, S., & Vo-Thanh, T. (2025). Understanding tourist barriers and personality influences in embracing generative AI for travel planning and decision-making. *International Journal of Hospitality Management*, *126*, 104105. https://doi.org/10.1016/j.ijhm.2025.104105

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton, Mifflin and Company.

Shaw, G., Bailey, A., & Williams, A. (2011). Aspects of service-dominant logic and its implications for tourism management: Examples from the hotel industry. *Tourism Management*, *32*(2), 207–214. https://doi.org/10.1016/j.tourman.2010.05.020

Sigala, M., Ooi, K.-B., Tan, G. W.-H., Aw, E. C.-X., Buhalis, D., Cham, T.-H., Chen, M.-M., Dwivedi, Y. K., Gretzel, U., Inversini, A., Jung, T., Law, R., & Ye, I. H. (2024). Understanding the impact of ChatGPT on tourism and hospitality: Trends, prospects and research agenda. *Journal of Hospitality and Tourism Management*, *60*, 384–390. https://doi.org/10.1016/j.jhtm.2024.08.004

Sparks, B. A., & Browning, V. (2011). The impact of online reviews on hotel booking intentions and perception of trust. *Tourism Management*, *32*(6), 1310–1323. https://doi.org/10.1016/j.tourman.2010.12.011

Sparks, B. A., So, K. K. F., & Bradley, G. L. (2016). Responding to negative online reviews: The effects of hotel responses on customer inferences of trust and concern. *Tourism Management*, *53*, 74–85. https://doi.org/10.1016/j.tourman.2015.09.011

Tan, K. P. S., Liu, Y. V., & Litvin, S. W. (2025). ChatGPT and online service recovery: How potential customers react to managerial responses of negative reviews. *Tourism Management*, *107*, 105057. https://doi.org/10.1016/j.tourman.2024.105057

Tetteh, M., & Thushara, M. (2023, May). Sentiment analysis tools for movie review evaluation-a survey. In *2023 7th international conference on intelligent computing and control systems (ICICCS)*, Madurai, India, 2023 (pp. 816–823). IEEE. https://doi.org/10.1109/ICICCS56967.2023.10142834

Umasuthan, H., Park, O. J., & Ryu, J. H. (2017). Influence of empathy on hotel guests' emotional service experience. *Journal of Services Marketing*, *31*(6), 618–635. https://doi.org/10.1108/JSM-06-2016-0220

Van Doorn, J., Lemon, K. N., Mittal, V., Nass, S., Pick, D., Pirner, P., & Verhoef, P. C. (2010). Customer engagement behavior: Theoretical foundations and research directions. *Journal of Service Research*, *13*(3), 253–266. https://doi.org/10.1177/1094670510375599

Van Thieu, N. (2024). Opfunu: An open-source python library for optimization benchmark functions. *Journal of Open Research Software*, *12*(1), 8. https://doi.org/10.5334/jors.508

Vargo, S. L., & Lusch, R. F. (2008). Service-dominant logic: Continuing the evolution. *Journal of the Academy of Marketing Science*, *36*(1), 1–10. https://doi.org/10.1007/s11747-007-0069-6

Vargo, S. L., & Lusch, R. F. (2014a). Evolving to a new dominant logic for marketing. In C. Vargo (Ed.), *The service-dominant logic of marketing* (pp. 3–28). Routledge.

Vargo, S. L., & Lusch, R. F. (2014b). Evolving to a new dominant logic for marketing. *Journal of Marketing*, *68*(1), 1–17. https://doi.org/10.1509/jmkg.68.1.1.24036

Xie, K. L., & So, K. K. F. (2018). The effects of reviewer expertise on future reputation, popularity, and financial performance of hotels: Insights from data-analytics. *Journal of Hospitality & Tourism Research*, *42*(8), 1187–1209. https://doi.org/10.1177/1096348017744016

Zeng, B., & Gerritsen, R. (2014). What do we know about social media in tourism? A review. *Tourism Management Perspectives*, *10*, 27–36. https://doi.org/10.1016/j.tmp.2014.01.001

Zhang, Z., Li, H., & Law, R. (2015). Differences and similarities in perceptions of hotel experience: The role of national cultures. *Journal of Travel & Tourism Marketing*, *32*(sup1), S2–S14. https://doi.org/10.1080/10548408.2014.959153

Zheng, Y., & Wu, P. F. (2022). Producing speed on demand: Reconfiguration of space and time in food delivery platform work. *Information Systems Journal*, *32*(5), 973–1004. https://doi.org/10.1111/isj.12377