Zhang H, Xiang Z, Zach FJ. Generative AI vs. humans in online hotel review management: A Task-Technology Fit perspective. *Tourism Management*.
DOI: 10.1016/j.tourman.2025.105187

Generative AI vs. Humans in Online Hotel Review Management: A Task-Technology Fit Perspective

Huihui Zhang[a*]
Assistant Professor
huihui.zhang@polyu.edu.hk
ORCID: 0000-0002-5206-1871

Zheng Xiang[b]
Professor
philxz@vt.edu
ORCID: 0000-0003-2608-4882

Florian J. Zach[b]
Associate Professor
florian@vt.edu
ORCID: 0000-0003-0243-4913

[a]School of Hotel and Tourism Management,
The Hong Kong Polytechnic University,
17 Science Museum Rd, Tsim Sha Tsui, Hong Kong

[b]Howard Feiertag Department of Hospitality and Tourism Management
Pamplin College of Business, Virginia Tech
295 West Campus Drive, 362 Wallace Hall
Blacksburg, VA 24061, USA

* Corresponding Author

Zhang H, Xiang Z, Zach FJ. Generative AI vs. humans in online hotel review management: A Task-Technology Fit perspective. *Tourism Management*.
DOI: 10.1016/j.tourman.2025.105187

Generative AI vs. Humans in Online Hotel Review Management: A Task-Technology Fit Perspective

Abstract

Despite Generative AI's ability to produce human-like content, its effectiveness as references for human responses, particularly in online review management, remains unclear. To address this question, this study explores if human responses resembling AI patterns are associated with enhanced customer perceptions. To provide deeper insights, we examined how this relationship shifts under varying technological and task conditions, guided by the Task-Technology Fit theory. In the empirical analysis, we automated responses to 32,129 online reviews using GPT, calculated the similarity between existing managerial responses and AI-generated content, and tested the relationship between human-AI similarity and the perceived helpfulness of review-response pairs. The findings reveal benefits of resembling AI with high model temperatures, particularly for positive reviews, while identifying negative outcomes under lower temperatures. This study enriches our understanding of an emerging technology that could have a huge impact on the industry and provides insights for practitioners to refine AI adoption strategies.

Keywords

Generative AI; Task-technology fit; Online reviews; Managerial response; Review valence; GPT; Big data analytics

1. Introduction

By January 2023, ChatGPT had attracted over 100 million users since its launch two months earlier, becoming the fastest-growing consumer internet application (Paris, 2023). This success spurred the prosperity of various Generative AI tools, such as Llama, Gemini, and Claude. These tools demonstrate strong capabilities in generating human-like content, such as text, images, and videos, showing transformative potential for value creation (Lohr, 2024). They could unlock trillions of dollars in value across multiple scenarios, such as product research and design, software engineering, marketing, and customer operations (McKinsey, 2023). In the hospitality and tourism industry, Generative AI is increasingly being integrated into various applications. For example, chatbots and virtual assistants built on this new technology have been deployed to offer personalized recommendations to improve how travelers plan their journeys (Ma, 2025; Marr, 2023). AI-generated content is also used to enhance customer online engagement by crafting dynamic and contextually relevant responses to social media interactions, inquiries, and online reviews (Touch, 2025).

Considering the promising productivity lift in customer operations, both practitioners and academics in diverse industries have started to explore the potential of Generative AI (Huang & Rust, 2024). As argued by recent studies, this early exploration stage is replete with opportunities but also concerns (Gursoy et al., 2023). The concerns include security and ethical issues, as well as the lack of empirical evidence demonstrating AI content's applicability to customer interactions. Therefore, tourism and hospitality firms remain hesitant to invest in this new technology (Mayer et al., 2025), although there are numerous tools specialized for domain-specific tasks, like online review response (Touch, 2025). However, at the individual level, content creators have actively embraced Generative AI for crafting personalized messages, such as responses to online reviews, and publishing them even with minor modifications (Microsoft, 2024). Therefore, it is important to explore whether AI-generated content is effective in customer engagement, which can serve as a basis for firms to establish further guidelines for individual or subsequent organizational adoption.

Several studies have explored the effectiveness of Generative AI in tourism-related applications (e.g., Morosan, 2025). However, these studies face limitations in the generalizability and validity of their findings. First, prior research has adopted a one-size-fits-all approach, treating Generative AI models as a uniform technology while overlooking the heterogeneity across different tasks and technological settings (Li & Lee, 2024; Morosan, 2025). Additionally, these studies heavily rely on primary data from interviews or surveys, where respondents are shown with a few examples of AI-generated responses (Shin et al., 2023; Tan et al., 2025). These approaches are insufficient to capture the dynamic nature of this new technology, as Generative AI can produce responses highly tailored to the input text, which can vary significantly due to the diversity of customer-generated content. Therefore, to yield more credible and generalizable findings, it is important to incorporate both task-specific and technological variations by leveraging big data analytics.

However, the major challenge for big data analytics lies in data availability before widespread real-world adoption. To address this issue, we propose an innovative approach that utilizes existing datasets of online reviews and managerial responses. Without having customers evaluating AI responses directly, we test if responses posted by hotel staff that are similar to AI-generated responses were perceived as more helpful. This design reflects the avenues available to humans when working with AI-generated content: adopt the AI text as-is, modify it, or even deviate entirely from the original text. By comparing human responses with high versus low similarity to AI-generated content, the results reveal whether AI response patterns align with customers' preferences.

For this study, we collected 32,129 TripAdvisor.com online reviews, each of which has received a response from hotel staff, in the Houston hotel market. Each online review was processed through GPT-3.5 model to generate an AI response. Next, we measured the similarity between the AI-generated responses and the existing managerial responses based on their embeddings. Regression analysis was applied to test the relationship between human-AI similarity and the perceived helpfulness of the review-response pairs. To account for variations across tasks and technological settings, this relationship was tested under varying conditions, including review sentiment and Generative AI temperature settings – a parameter that decides the variability and creativity of generated responses. These conditions were introduced because the effectiveness of technology is dependent on whether the technology fits the specific task requirements, as guided by the Task-Technology Fit (TTF) theory. The results show that, when displaying a high similarity to AI content generated under the default mode of lower temperatures, human responses are likely to be perceived as less helpful. In contrast, high similarity to AI content generated under high model temperatures tends to enhance the perceived helpfulness of human responses, especially for positive reviews.

The findings reveal that fully following AI content generated under default temperature settings fails to meet the customers' preferences, highlighting the risks of a wide range of individual usage without professional training and guidance in customer engagement settings. This urges practitioners to establish AI adoption guidelines for their employees to maintain the quality of customer service, offering invaluable practical insights. Additionally, this study brings theoretical contributions to the hospitality and tourism literature, by introducing TTF theory to understand the optimal adoption of Generative AI, emphasizing the alignment between technology settings and task-specific requirements.

2. Literature Review

2.1 Generative AI Applications in Hospitality and Tourism

Artificial Intelligence (AI) is defined as a system or machine that thinks and acts humanly, or thinks and acts rationally (Russell & Norvig, 2010). Generative AI refers to a subset of AI capable of creating new content such as text, images, and videos. A typical example is ChatGPT, a natural language generation model powered by generative pre-trained transformers (GPT) (Van Dis et al., 2023). These transformer-based Generative AI models exhibit unprecedented intelligence, allowing them to understand human languages and generate content that is even indistinguishable from human writing (Liu et al., 2023). This ability is posited to revolutionize content automation that requires human knowledge and creativity (Reisenbichler et al., 2022). In many domains, such as programming and medical diagnosis, Generative AI models have been identified to match and even outperform humans (Eriksen et al., 2024; Hou & Ji, 2024). However, unlike questions in natural science fields with definitive answers, tasks in social science domains such as tourism are more complex due to the subjective nature of human interactions. This makes evaluating the effectiveness of Generative AI in such contexts a challenging task.

In hospitality and tourism literature, many studies have explored this emerging technology, but most of them focus on conceptual perspectives to discuss the potential benefits and risks associated with Generative AI (e.g., Gursoy et al., 2023). However, there is limited empirical evidence on its effectiveness in addressing human-centered interactions. To address this gap, recent studies have employed experiments and surveys to explore the applications of Generative AI across various traveler-centered scenarios. For example, several studies have investigated Generative AI's role in assisting decision-making, building trust, and enhancing customer perceptions and behavioral intentions in travel planning (Kim et al., 2023; Shin et al., 2023). Others have examined its impacts in the context of customer engagement, particularly in online review response (Litvin & Tan, 2024; Tan et al., 2025).

Although prior empirical research has provided valuable insights, it fails to capture the dynamism and complexity of Generative AI, as well as its level of intelligence. Previous studies either rely on a few responses to evaluate customers' reactions (Shin et al., 2023; Tan et al., 2025), or treat Generative AI as a generic tool without showing any specific cases, asking only for users' perceptions based on their own use experiences (Li & Lee, 2024; Morosan, 2025). This approach largely limits the credibility and generalizability of the findings, especially given the unique nature of Generative AI's output.

Generative AI's output is unstructured, often comprising qualitative content like text or video, whose quality is inherently difficult to assess (Hu & Zhou, 2024). Moreover, its output exhibits significant variability. One reason is that Generative AI produces responses highly tailored to the input, such as online reviews (Van Dis et al., 2023). Since these inputs are often heterogeneous, the resulting outputs can vary substantially (Heya et al., 2024). Therefore, the difficulties in assessing the quality of Generative AI's output combined with its significant variability highlight the inadequacy of survey- or interview-based methods that rely on general perceptions or a limited number of examples. This underscores the importance of adopting large-scale data to understand the effectiveness of this unique technology.

In addition, prior literature focusing on Generative AI's application in the hospitality and tourism industry has predominantly adopted a one-size-fits-all approach, further limiting their generalizability. Most previous studies tested Generative AI's capabilities without considering conditional factors (Li & Lee, 2024; Morosan, 2025), such as variations in technological and task settings. However, the configuration of technologies largely influences their applicability to specific contexts (Lu et al., 2021). From the task perspective, only a few studies have embraced the characteristics of applied tasks, such as the size of initial options for option reduction task (Shin et al., 2023) and type of listing on Airbnb promotion tasks (Choi et al., 2024). These studies suggested that customers hold heterogeneous expectations, which influence the effectiveness of Generative AI during the interactions. Given the importance of both task-specific and technological features, it is important to adopt a more nuanced approach to evaluate the performance and applicability of Generative AI across diverse scenarios.

2.2 Online Review Responses

This study focuses on managerial responses to online reviews given its relevance and importance to publicly display welcoming customer feedback and to engage customers through personalized communication (Tan et al., 2025; Zhang et al., 2024). First, understanding the application of Generative AI to craft online review responses highly aligns with the strength of this new technology. Generative AI models demonstrate extraordinary capabilities in creating textual content. Previous studies identified that Generative AI models (e.g., GPT) can pass Turing test in many scenarios, including online review response (Tan et al., 2025), which means customers cannot distinguish between AI- vs human-generated textual content.

Second, responding to online reviews is increasingly important in this digital age. Consumers rely heavily on online reviews in decision-making, especially when purchasing intangible products like hospitality and tourism services (Xiang & Gretzel, 2010). Managing online reviews has become an essential component of customer relationship management because reviews can significantly influence consumer perceptions and behaviors, and directly impact financial performance (Nicolau et al., 2024). Therefore, many firms have been proactively engaging with customers through managerial responses, facilitating two-way communications. However, only around 30 percent of hotel reviews receive managerial responses typically due to labor shortages (Kwok, 2022). As Generative AI is powerful to at least draft responses, it is important to explore it as a viable alternative solution to this dilemma. For example, several software providers are already offering response drafts to online reviews (Touch, 2025).

Online review management is a difficult task, not just for Generative AI, but already persistently challenging for humans. Therefore, online reviews and managerial responses have been widely studied in multiple industries, especially in the hotel and tourism sector which involves intense human interactions. Previous studies have identified that firms need to adapt their responding strategies with diverse online review characteristics, such as review valence. To fit different expectations from satisfied versus dissatisfied customers, firms are recommended to adjust the style of responses, spanning multiple textual features such as inter-response variation, relevance to review texts, and lexical diversity of response (Lopes et al., 2023).

Dissatisfied customers place greater importance on inter-response variation compared to satisfied customers. They perceive customized responses as a reflection of attentiveness from service providers, while templated responses often fail to convey such care (Liu et al., 2021; Zhang et al., 2019). Therefore, when dealing with negative reviews, high repetitiveness in responses can intensify consumer dissatisfaction (Wei et al., 2013; Zhang et al., 2024). However, this effective is less pronounced for positive reviews (Shin et al., 2020). There are even studies indicating no significant difference between templated and personalized responses to positive reviews while it matters for negative reviews (Wei et al., 2013).

There are several optional approaches to tailor managerial responses, such as responding directly to the specific aspects mentioned in the original reviews. Responding by paraphrasing review texts can foster positive customer relations because these responses serve as verbal cues of active listening, reflecting firms' empathy and genuine care for the customer's voice (Wang & Jia, 2023). This is particularly important for negative reviews, where customers anticipate detailed apologies, explanations, and accommodative actions in response to service failures (Min et al., 2015; Zhang et al., 2020). On the contrary, paraphrasing positive reviews may lead to weaker or even adverse outcomes, as highly specific responses to praises encounter the risk of overselling and may trigger customers' psychological reactance (Deng & Ravichandran, 2023; K. L. Xie et al., 2017).

Response strategies also vary within a single reply, regarding lexical diversity. A response can be formed by diverse words or more uniform terms. Xu and Zhao (2022) found that low word diversity is beneficial for responding to negative reviews as using a smaller set of words shows a focused orientation, targeting key points reflected in the complaints. However, this response strategy is not significant for positive reviews. The heterogeneous requirements for response strategies based on review valence further highlight the limitations of the "one-size-fits-all" approach in understanding Generative AI adoption for online review responses.

3. Research Design and Hypotheses

The dynamic nature of Generative AI and the complexity of online review response strategies make it challenging to examine the effectiveness of this new technology. Task-Technology Fit (TTF) theory offers a valuable framework to address this challenge by emphasizing the alignment between technological characteristics and task-specific requirements (Goodhue & Thompson, 1995). This theory has been extensively examined across multiple fields, including in Information Systems (IS), marketing, and tourism (Lin et al., 2020; Xu et al., 2024). The following framework derived from TTF is developed to guide this study (Figure 1).

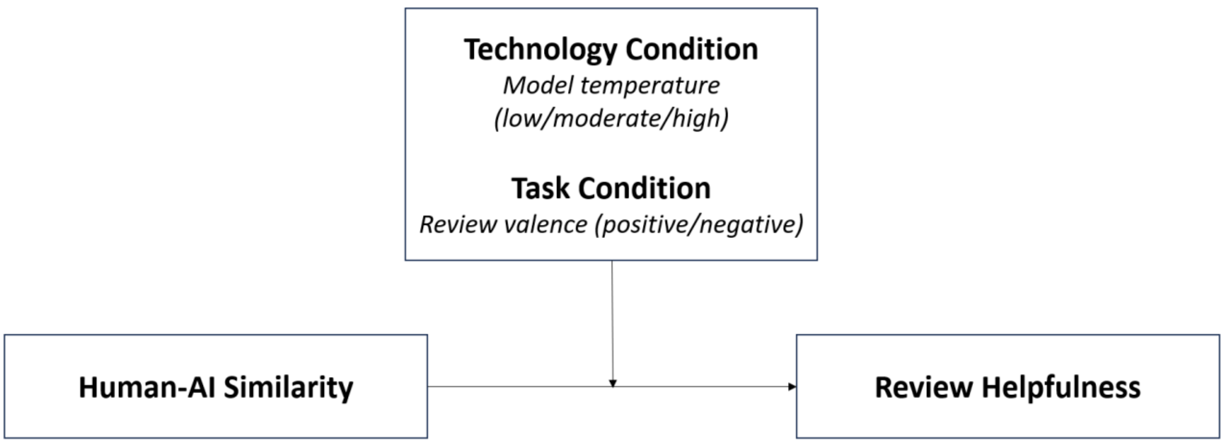

Figure 1. Research framework

Employing the TTF framework, previous studies have shown that performance is enhanced when the applied technology fits the task (Shin & Jeong, 2022). We argue that, in general, Generative AI can meet the demands of online review response. This capability arises from its advanced design. First, these models are typically trained on extensive data sets with trillions of parameters (Brown et al., 2020). For instance, GPT draws knowledge from thousands of books, millions of web pages, and other diverse materials, enabling it to comprehend and synthesize information across multiple domains, such as marketing, psychology, and communication (Yang et al., 2023). This guarantees the quality of generated content. Second, the next-word prediction algorithm keeps responses staying highly relevant to the original text (Bansal et al., 2024). Third, machine learning ensures AI-generated responses are consistently logical, unlike some human responses which might be influenced by emotions or biases (Liang et al., 2021). The quality, relevance, and rationality of AI-generated content make it a reliable template for hotel staff when addressing online reviews. As such, we argue that AI-generated content is effective enough to meet customers' demands, which is indicated by a positive relationship between human-AI similarity and perceived helpfulness.

Notably, this research design did not have customers read the AI-generated responses and report their perceptions, considering the challenges and costs of collecting direct customer feedback on large-scale AI-generated content. Therefore, we applied an indirect approach to assess the effectiveness of AI content by leveraging historical human responses and their associated evaluations from real-world customers. Specifically, we examined whether human responses that closely resembled AI responses were perceived as more helpful. This design is grounded in the practice of typical AI-adoption in content creation, wherein individuals use AI-generated content as a basis to craft tailored text. That is, AI-generated content can become the norm that humans will seek to emulate and adapt to specific use cases. If we can identify human content that is similar to this new norm is associated with enhanced customer perceptions, it underscores the value of AI-generated content. Therefore, we hypothesize that:

H1: Human responses with high similarity to AI-generated content are likely to be perceived as more helpful.

The value of TTF framework lies not only in its focus on understanding how the general alignment between technology and tasks influences task performance, but also in its advocacy for deconstructing task and technology to examine how the configuration of different features facilitates optimal outcomes. This approach has been applied to understand the adoption of various technologies (Lin et al., 2020; Xu et al., 2024). The effectiveness of these technologies may vary with the characteristics of tasks. For example, a recent study on Generative AI adoption in marketing for chewing gum manufacturers also concluded that its utility varies across organizational tasks, with greater benefits expected in tasks focused on automation and innovation (Przegalinska et al., 2025). When responding to online reviews, the effectiveness of this new technology may vary with review polarity. For instance, Litvin and Tan (2024) identified that AI-generated responses are favored by readers in general, while Tan et al. (2025) found that Generative AI responses to negative reviews result in lower affective, cognitive, and conative outcomes. This divergence in results might stem from the varying expectations of readers when evaluating responses to negative and positive reviews. It highlights the importance of exploring how the effectiveness of Generative AI varies with review valence to identify if it explains the varied outcomes.

In addition to task characteristics, the effectiveness of technology may also vary with technology features (Itani et al., 2020; Rivera et al., 2016; Zhou et al., 2020). For example, Rivera et al. (2016) found that mobile services provided on smart phones with location awareness and mapping options offer higher utilities than those on tablets without such functionalities. Similarly, Itani et al. (2020) and Zhou et al. (2020) identified the influence of technology features when adopting social media technologies. However, prior research regarding Generative AI overlooks the varying configurations of this technology, oversimplifying its adaptability. This limits our understanding of the optimal implementation of Generative AI. To fill this gap, this study evaluates the effectiveness of Generative AI under varying task and technology conditions, as shown as the moderating variables in Figure 1.

This research focuses on model temperature because it is a crucial parameter that has attracted significant attention across multiple domains, yet determining its optimal setting remains a challenge. Temperature is a crucial hyperparameter that adjusts the randomness of generated text (Ziegler et al., 2019). Higher temperatures introduce more variation and creativity by allowing the model to select lower-probability words, while low temperatures lead to more deterministic and conservative outputs (Bhavya et al., 2022). For example, at a low temperature, a Generative AI model might respond to a positive review with words that are highly possible to appear in this context, such as “thank you” and “appreciate”, leading to very conventional responses. In contrast, higher temperatures allow for the usage of less frequent words like “serendipity” and “motivation”, thus increasing the variability of responses.

Previous studies have explored the optimal temperature across multiple tasks. A lower temperature is preferred in natural science fields such as chemistry and radiography (Mukherjee et al., 2023). High temperatures, however, yield better performance in tasks like analogy and story creation (Lucy & Bamman, 2021). For tasks related to common sense or human intent, which require both logical reasoning and flexible resonance, a moderate level of temperature might excel (Huang et al., 2022). However, some studies involving social science tasks identified an insignificant influence of temperature (Koneru et al., 2023). As a task fusing knowledge understanding and human interactions, online review management presents a challenge in selecting the optimal temperature.

We argue that the effectiveness of Generative AI depends on how well this technology, under different temperature settings, fits the specific needs of responding to positive and negative customer feedback, as summarized in Table 1. For negative reviews, there are rationales for both high and low temperatures. Considering response variability, high temperatures are more suitable because they lead to more diverse outputs (Mukherjee et al., 2023), which are needed in responding to complaints (Wei et al., 2013). However, low temperatures might be preferred because they tend to generate responses more relevant to the original reviews with more concentrated vocabulary. The direct and specific replies are effective in addressing negative reviews (Xu & Zhao, 2022; Zhang et al., 2020). As there are more conditions requiring lower model temperatures when responding to positive reviews, we hypothesize that:

H2a: For negative reviews, lower model temperatures enhance the positive association between human-AI similarity and perceived helpfulness.

For positive reviews, high temperatures are favored because they tend to generate responses less relevant to the inputs, which will help avoid the overselling risk (Ravichandran & Deng, 2023). In addition, high temperatures introduce more variations across drafted responses, which sometimes enhance customer perceptions of positive reviews (Shin et al., 2020). Therefore, we raise the following hypothesis:

H2b: For positive reviews, higher model temperatures enhance the positive association between human-AI similarity and perceived helpfulness.

Table 1. Conceptual match between task feature and technological characteristic

Response feature	**Negative reviews**	
	Task requirement	Preferred temperature
Variation across response	High (Shin et al., 2020; Wei et al., 2013)	High
Review-response relevance	High (Min et al., 2015; Zhang et al., 2020)	Low
Lexical diversity	Low (Xu & Zhao, 2022)	Low
Response feature	**Positive reviews**	
	Task requirement	Preferred temperature
Variation across response	High (Shin et al., 2020) or insignificant (Wei et al., 2013)	High
Review-response relevance	Low (Deng & Ravichandran, 2023; Zhang et al., 2020)	High
Lexical diversity	Insignificant (Xu & Zhao, 2022)	——

4. Methodology

To examine the effectiveness of Generative AI in response to positive versus negative reviews under three temperature settings, we conducted an empirical analysis outlined in Figure 2. It begins with collecting data from TripAdvisor.com, followed by AI response generation, human-AI comparison, and regression. More detailed information on these stages will be illustrated in the following subsections.

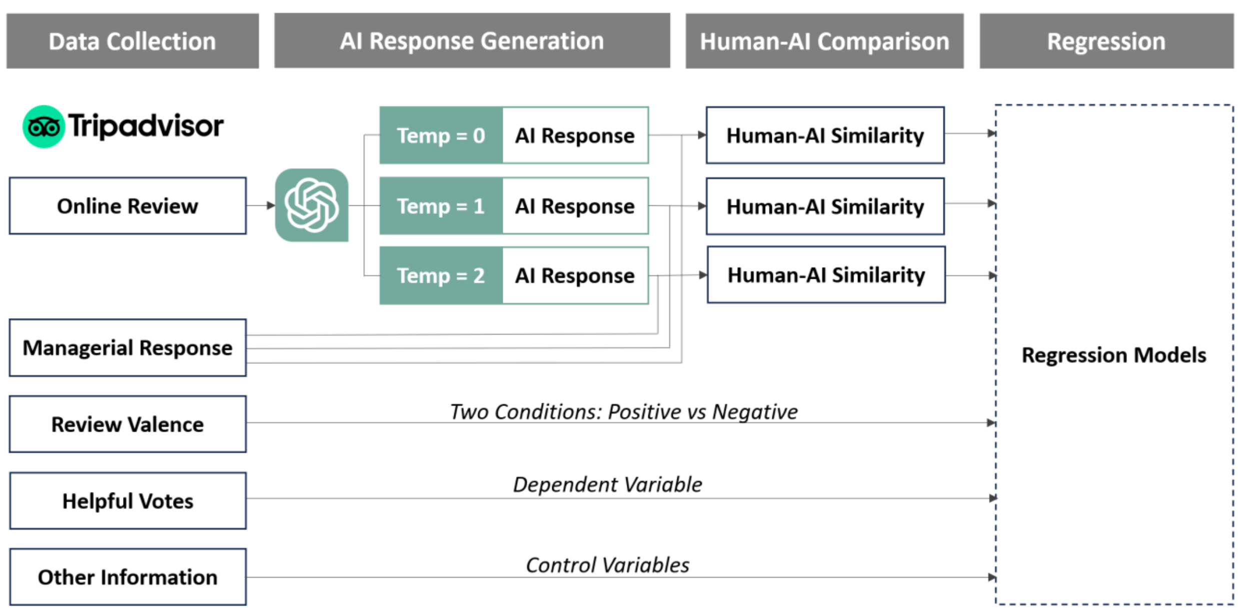

Figure 2. Process of data analysis

4.1 Data Collection

This study employs online reviews and corresponding managerial responses collected in September 2023 from TripAdvisor.com, a social media platform widely used in previous studies (e.g., K. Xie et al., 2017). Our dataset comprises English reviews posted for hotels in Houston, Texas, covering key information on review content, managerial response, review valence, helpful votes, and other details about the hotel and reviewer (Figure 3). As this study focuses on the similarity of managerial response to AI generated content, only reviews that have received managerial responses are retained. After removing observations with missing values, 32,129 observations are kept for further analysis. These observations are composed of 28,397 positive reviews (with a rating of four or five) and 3,742 negative reviews (with a rating of one or two). 168 unique hotels are covered, and the distribution of hotel star rating is shown in Figure 4.

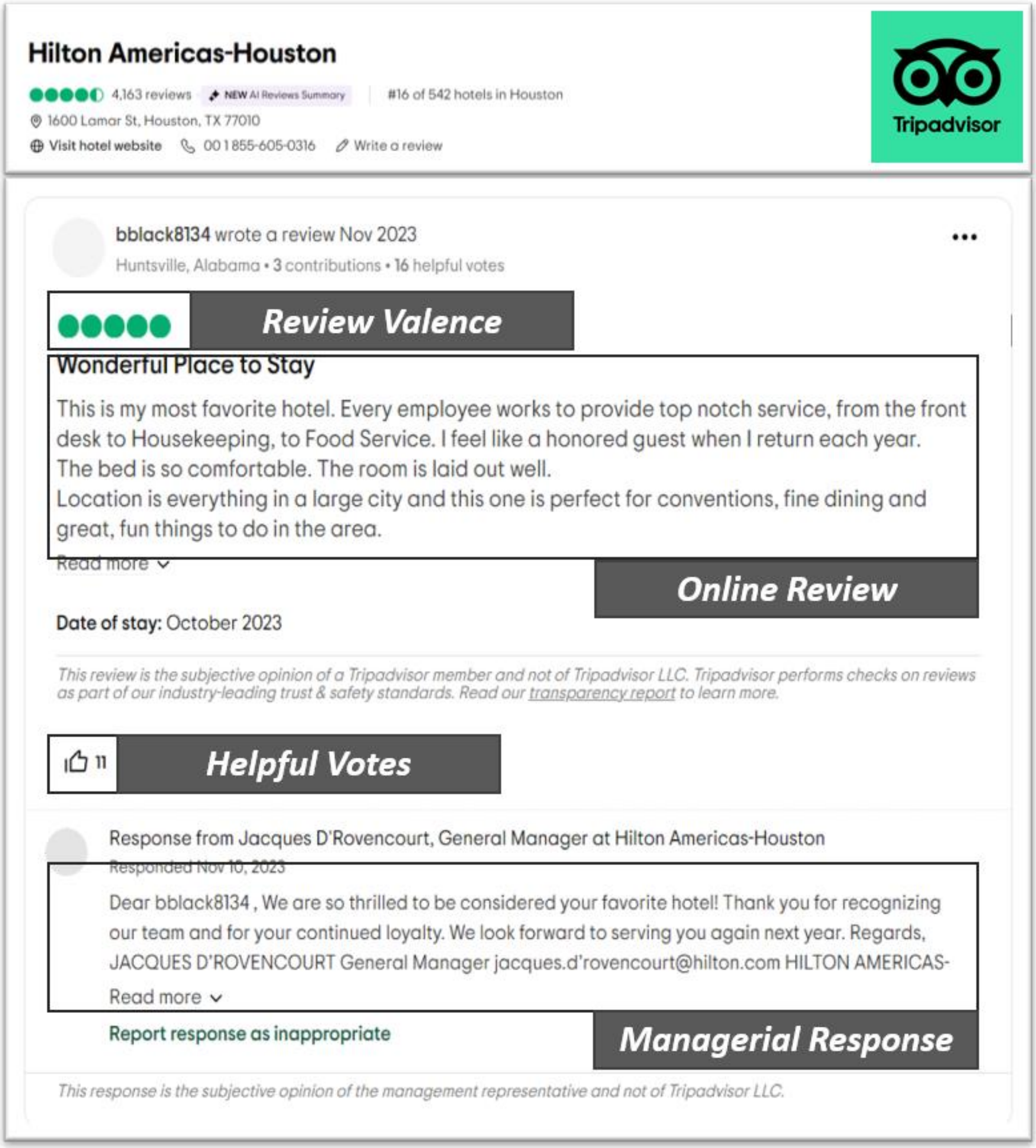

Figure 3. Example of online review information collected from TripAdvisor.com

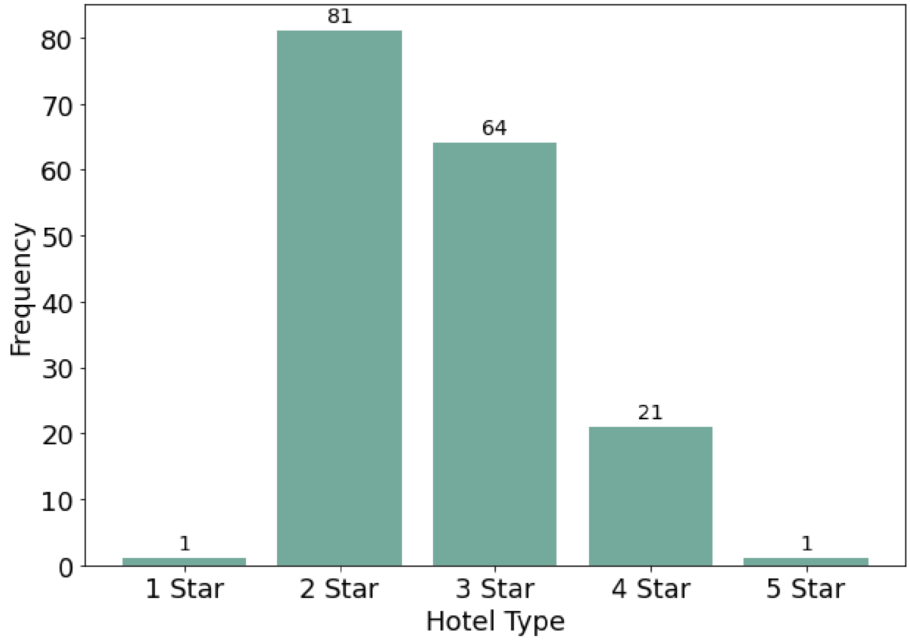

Figure 4. Distribution of hotel star rating

4.2 AI Response Generation

This study utilizes Python to access GPT-3.5 Turbo models through the chat completion component of the OpenAI API to generate AI responses to online reviews (Reisenbichler et al., 2022). To evaluate the effectiveness of Generative AI under varying technological settings, we conducted the generation process three times, setting the ***temperature*** parameter to 0, 1, and 2. These temperatures represent the ***Conservative*** (Temp = 0), ***Default*** (Temp = 1), and ***Creative*** (Temp = 2) modes of GPT-3.5, with 0 and 2 being the extremes and 1 the default setting, which is accessible to users with subscriptions. This approach allows us to test Generative AI's performance across a range of output styles, from highly predictable to more varied and creative. Typical examples of AI responses under different temperature settings are displayed in Appendix A.

During each round of generation, every online review is processed individually, so that the GPT model can craft a tailored response to it. After the generation process, each online review is paired with one managerial response posted by hotel staff and one AI response generated by the GPT model. The prompt used for response generation is formed accordingly by concatenating online review title and content to a standard instruction as shown in Figure 5. The reason why we expect the model to generate responses of around 65 words is because it is the average length of managerial responses in our dataset. A similar length distribution will make the responses from different sources, i.e., hotel staff and GPT, more comparable by avoiding distractions caused by significant differences in response length.

Zhang H, Xiang Z, Zach FJ. Generative AI vs. humans in online hotel review management: A Task-Technology Fit perspective. *Tourism Management*.
DOI: 10.1016/j.tourman.2025.105187

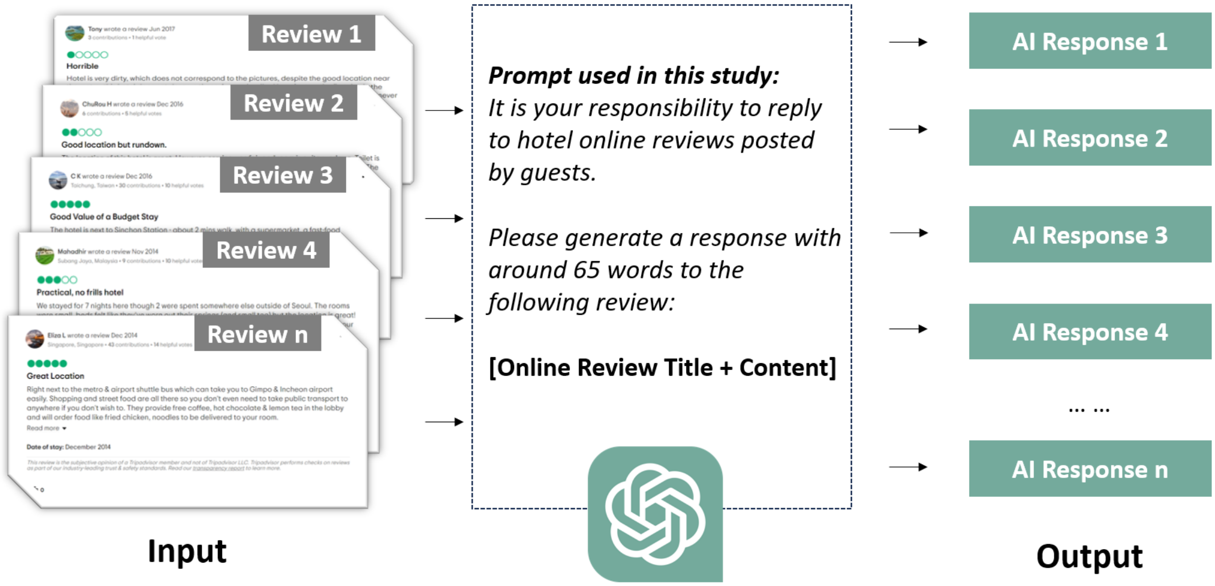

Figure 5. Process of AI response generation

4.3 Variables

Perceived ***helpfulness*** is used to operationalize the effectiveness of responses, quantified by the number of helpful votes received by an online review-response pair. This proxy is selected for the following reasons. First, it reflects the attitudes of customers toward overall communication involving both the original review text and the managerial response. This measure is directly correlated with the effectiveness of review-response interactions (Kwok & Xie, 2016), thus it can serve as a proxy for the response quality. It has been widely employed in previous studies for examining the effectiveness of response strategies (e.g., Li et al., 2017). Second, this measure allows us to conduct this study at the review level, providing insights into how different types of reviews (i.e., positive vs negative) benefit asymmetrically from Generative AI. Additionally, to address potential bias from reviews receiving helpful votes before a response is posted, we incorporate response time lag as a control variable to better account for the effect of managerial responses on helpfulness perceptions.

The similarity between human responses and AI responses is calculated based on their embeddings (Carlson, 2023). Text embedding refers to a numerical representation of a textual message, which contains meaningful semantic information. Specifically, for each online review, its managerial response and AI response are converted into embeddings using Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2018), a powerful method for computing document embeddings that has been widely used in business tasks (e.g., Carlson, 2023). This method is advantageous over traditional approaches such as Doc2Vec because it captures the contextual meaning of each word, which is particularly important in understanding the nuances of human communication. The degree of ***human-AI similarity*** is then quantified by the cosine similarity between the response embeddings (Figure 6).

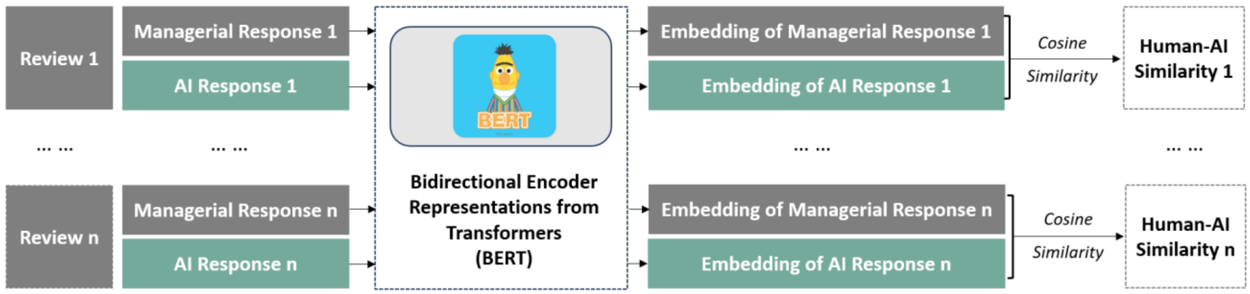

Figure 6. Measurement of human-AI similarity

This study also captures how task characteristics shift the effectiveness of Generative AI by taking ***review valence*** as a moderator. Review valence is measured by a binary variable which takes reviews with ratings of one or two as negative and those with ratings of four or five as positive. To ensure the validity and reliability of research findings, we also incorporate multiple control variables in the following regression analysis. These control variables include review features, such as title length, content length, readability, photo, recency, and timeliness. We further control reviewer characteristics (i.e., reviewer expertise), response features (i.e., length, readability, timeliness, and respondent level), and hotel attributes (i.e., overall rating, review volume, class, and ranking). These factors are identified as key determinants of review helpfulness (e.g., Shin et al., 2021). Detailed descriptions and statistics of these variables are displayed in Tables 2 and 3. The correlation table can be found in Appendix B. It is important to note that ***human-AI similarity*** is measured separately for each temperature setting (Temp = 0, 1 or 2), as this variable varies with temperature. Meanwhile, all the other variables for each review remain constant.

To obtain a more straightforward comparison between humans and AI models, we transform high-dimensional text embeddings into a two-dimensional space with a dimensionality reduction technique, t-Distributed Stochastic Neighbor Embedding (t-SNE) (Van der Maaten & Hinton, 2008). One major reason for choosing t-SNE is its ability to handle non-linear relationships in the data, which is common for textual information. Another reason is its superior capability to capture the essential similarities and distinctions between different types of embeddings (Zhou et al., 2023). We feed four sets of embeddings: human responses and AI responses from three different temperatures to the t-SNE model. The results are visualized and discussed in Section 5.

4.4 Regression Model Specification

As helpfulness is a count variable that exhibits overdispersion, following Zhang et al. (2024), this study employs a negative binomial regression method. The regression models are estimated with the following equation, where β refers to the estimated coefficient and ε_i denotes the error terms. The first equation identifies the main effect of human-AI similarity, while the second one reveals the moderating effects of review valence.

Zhang H, Xiang Z, Zach FJ. Generative AI vs. humans in online hotel review management: A Task-Technology Fit perspective. *Tourism Management*.
DOI: 10.1016/j.tourman.2025.105187

$$Helpfulness_i = \beta_1 * HumanAISim_i + \beta_2 * Controls_i + \varepsilon_i$$

$$Helpfulness_i = \beta_1 * HumanAISim_i + \beta_2 * HumanAISim_i * ReviewValence_i + \beta_3 * Controls_i + \varepsilon_i$$

To understand how technological characteristics shift the impact of human-AI similarity, we estimate these regression models under each temperature setting (Temp = 0, 1, or 2).

Table 2. Variable summary

Variable name	**Variable description**
Helpfulness	Number of helpful votes received by an online review
Human-AI similarity	Managerial response's similarity to AI generated content responding to the same online review
Review valence	A categorical variable indicating if an online review is positive or negative
Review title length	Number of words in the title of an online review
Review content length	Number of words in the content of an online review
Review readability	Gunning fog readability index of the content of an online review
Review photo	A binary variable indicating if an online review is posted with photos
Review recency	Number of months since an online review was posted
Review time lag	Number of months between the date when an online review was posted and the date of stay
Reviewer expertise	Number of reviews that have been posted by a reviewer
Response length	Number of words in a managerial response
Response readability	Gunning fog readability index of a managerial response
Response sentiment	Sentiment polarity of a managerial response
Response time lag	Number of months between the date when an online review was posted and the date when its managerial response was posted
Respondent level	A categorical variable indicating the level of respondent who generated a managerial response: Administrative, Executive, and Operational
Hotel review rating	Average rating of online reviews received by a hotel
Hotel review volume	Number of online reviews received by a hotel
Hotel class	Hotel scales: 5 - luxury hotel, 4 - above-average hotel, 3 - full-service hotel, 2 - mid-market economy hotel, 1 - budget traveler hotel
Hotel ranking	Ranking of a hotel among all hotels from the same city

Table 3. Variable statistics

Variable	Mean	Std. Dev.	Min	Max
Helpfulness	0.10	0.75	0	76
Human-AI similarity (Temp = 0)	0.88	0.05	0.61	0.98
Human-AI similarity (Temp = 1)	0.88	0.05	0.61	0.98
Human-AI similarity (Temp = 2)	0.79	0.09	0.10	0.98
Review valence – Negative	0.12	——	0	1
Review title length	4.15	2.74	0	30
Review content length	78.97	43.49	5	1258
Review readability	8.38	3.31	1.76	97.11
Review photo	0.03	0.17	0	1
Review recency	85.94	37.23	8	252
Review time lag	0.43	1.37	0	53
Reviewer expertise	80.33	579.94	0	70,083
Response length	60.76	31.75	1	412
Response readability	8.72	2.49	0.40	44.66
Response sentiment	0.35	0.19	-0.82	1
Response time lag	10.33	34.08	0	236
Respondent level – Administrative	0.75	——	0	1
Respondent level – Executive	0.02	——	0	1
Hotel review rating	4.28	0.44	1	5
Hotel review volume	1742.34	1291.74	1	4065
Hotel class	3.47	0.71	1	5
Hotel ranking	69.58	85.99	1	532

5. Results and Discussion

After reducing the dimensionality of human and AI text embeddings, we visualize the outputs of t-SNE in Figures 7-9. The two axes do not have specific meanings but serve as tools to reflect the similarities and differences between humans and AI. Figure 7 presents the embeddings from human responses compared to AI responses generated at the lowest model temperature setting (Temp=0). It shows a significant overlap between human and AI-generated responses. Figure 8 shows a similar pattern, but with slightly more dispersion among AI-generated responses, when the model temperature is set to 1. The alignment between human responses and AI texts under lower temperatures reflects AI's ability to produce relevant and reasonable responses, closely emulating human-like interactions. The highest temperature setting (Temp=2) shows the most distinct patterns, with AI responses forming a separate cluster from human responses (Figure 9). This indicates that high temperatures lead to greater variability in AI responses, which may result in outputs that deviate significantly from typical human responses. While this could be advantageous for applications due to more creative expressions, it also risks reducing the contextual appropriateness of the AI-generated text.

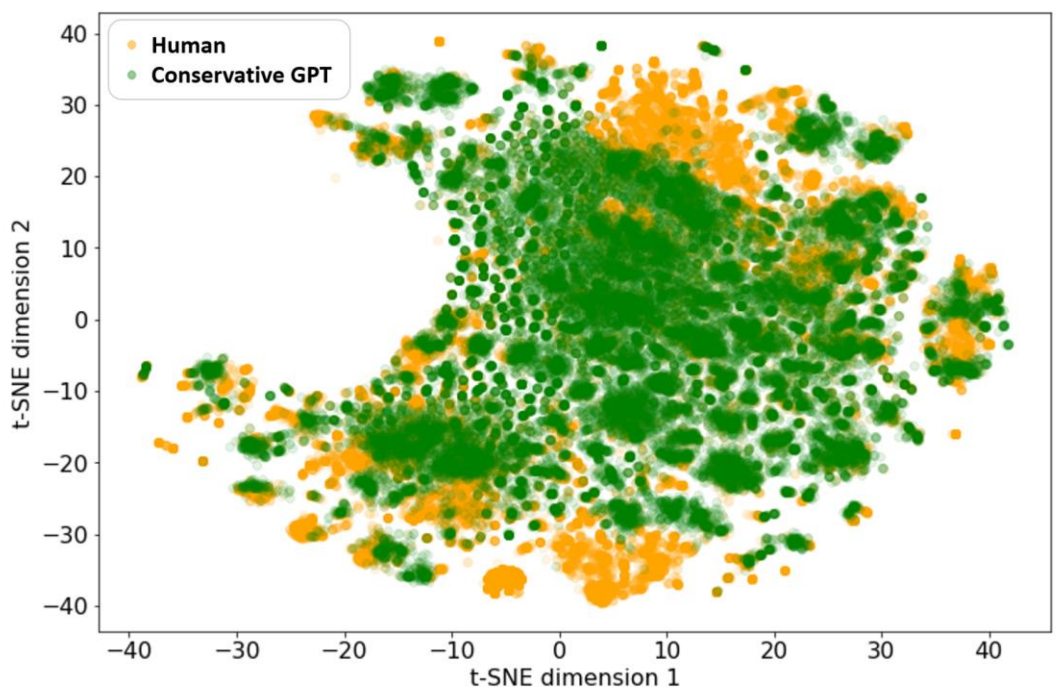

Figure 7. Text embedding comparison between human and AI (Temp=0)

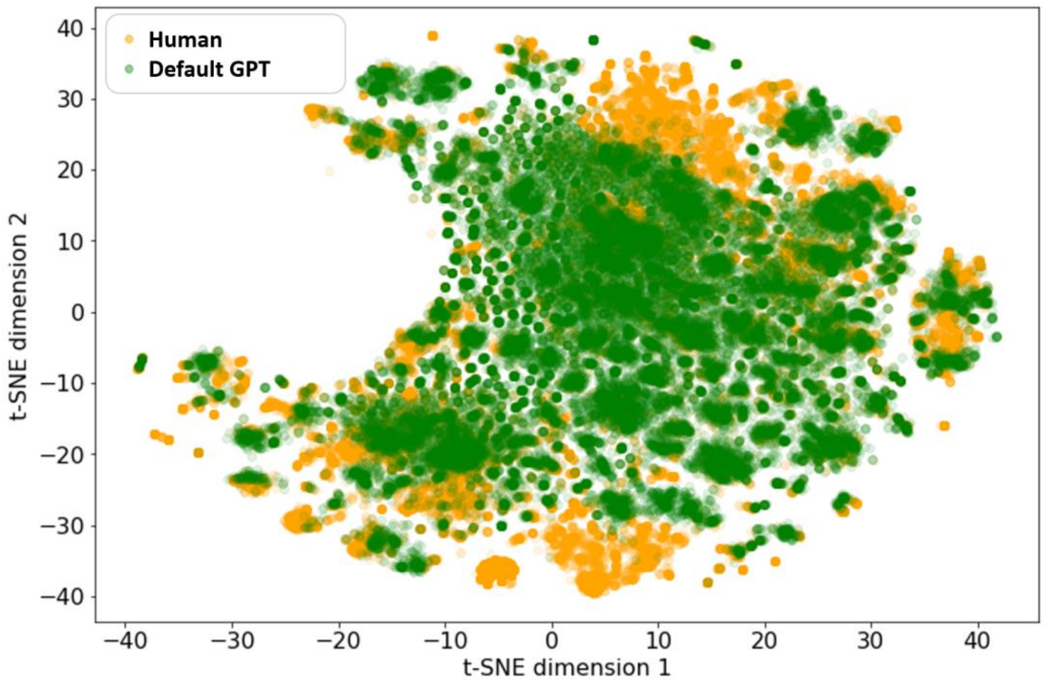

Figure 8. Text embedding comparison between human and AI (Temp=1)

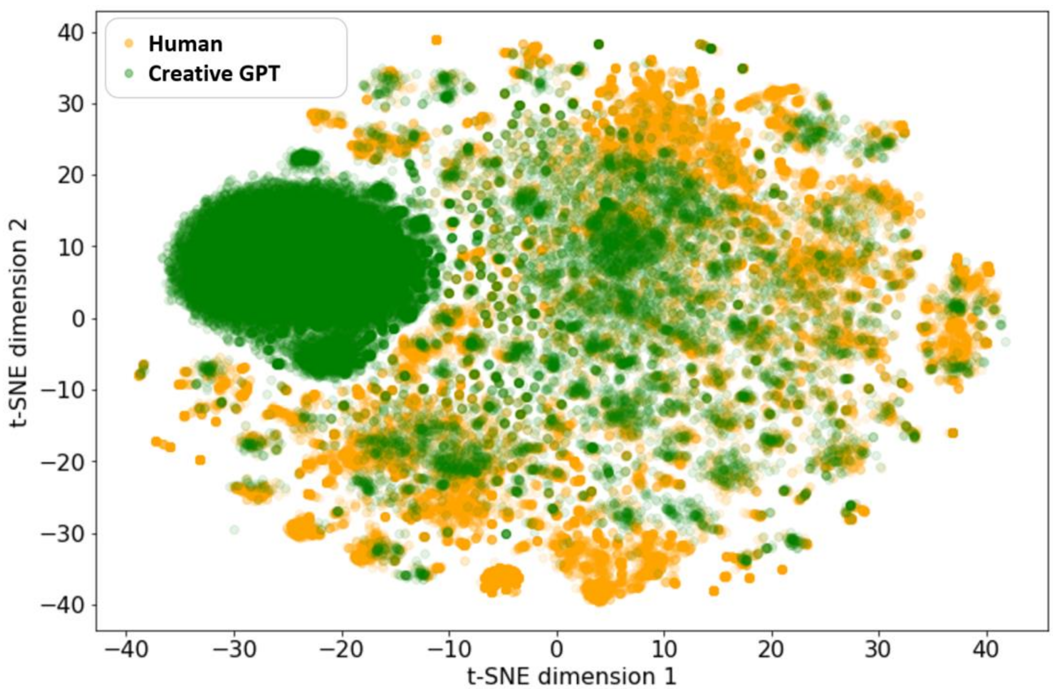

Figure 9. Text embedding comparison between human and AI (Temp=2)

As argued in Section 2.3, the adjustment of temperature will lead to changes in the drafted responses concerning three aspects: variation across responses, relevance to input review, and lexical diversity within one response. We compare these features of the content produced by AI models with different temperature settings and hotel staff. When quantifying the content features, variation across responses is measured by comparing one response to others for the most recent five reviews from the same hotels following Zhang et al. (2019). The relevance of a response to its corresponding review is represented by the Jaccard similarity which compares their overlapped terms against the union of used terms (Wang & Jia, 2023). Following Xu and Zhao (2022), we measure lexical diversity by the ratio of unique words to the total words in a response.

Figures 10 to 12 display the comparison in content features across AI models and humans. Models with high temperatures (Temp =2) yield responses with higher variability, lower relevance to input (i.e., original review text), and high lexical diversity. This finding is reasonable because high temperatures increase the randomness in generated texts, leading to a broader range of word choice and response framing (Mukherjee et al., 2023). The randomness and flexibility in the generated content also reduces responses' coherence to the input texts (Bhavya et al., 2022). At the same time, low and medium temperatures (Temp = 0 or 1) show opposite patterns. These findings align with previous research in natural science domains (Mukherjee et al., 2023). It suggests that models with high temperatures allow more flexibility, indicating a more creative model, while low temperatures prioritize predictability, reflecting a conservative orientation.

In terms of human responses, the content generated by hotel staff displays a moderate level of response variability, review-response relevance, and lexical diversity, between the highest and lowest range of AI models. Notably, human responses display a significant difference between negative and positive reviews, reflecting the strategies suggested in previous studies (Deng & Ravichandran, 2023; Wei et al., 2013; Xu & Zhao, 2022). For negative reviews, hotel staff tend to generate more specific responses with higher relevance to the original review with concentrated word choice. The alignment between practice and academic knowledge implies that hotel staff may adapt their operations according to their experience or cumulated knowledge, signaling potential human learning behavior. However, this pattern is absent from the AI-generated content studied in this research.

Zhang H, Xiang Z, Zach FJ. Generative AI vs. humans in online hotel review management: A Task-Technology Fit perspective. *Tourism Management*.
DOI: 10.1016/j.tourman.2025.105187

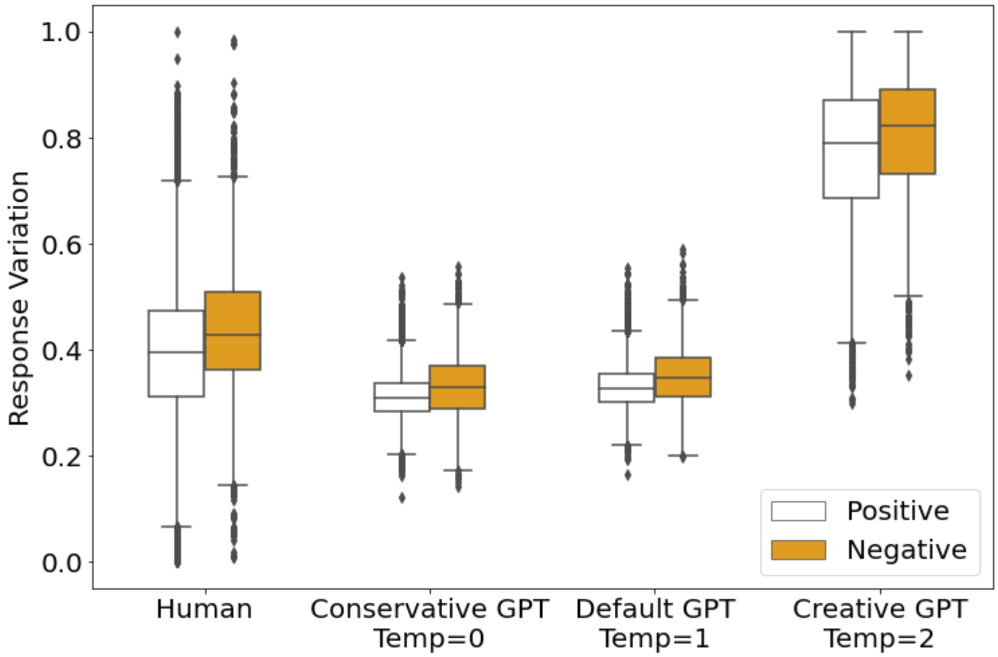

Figure 10. Degree of response variation from the same hotel across different sources

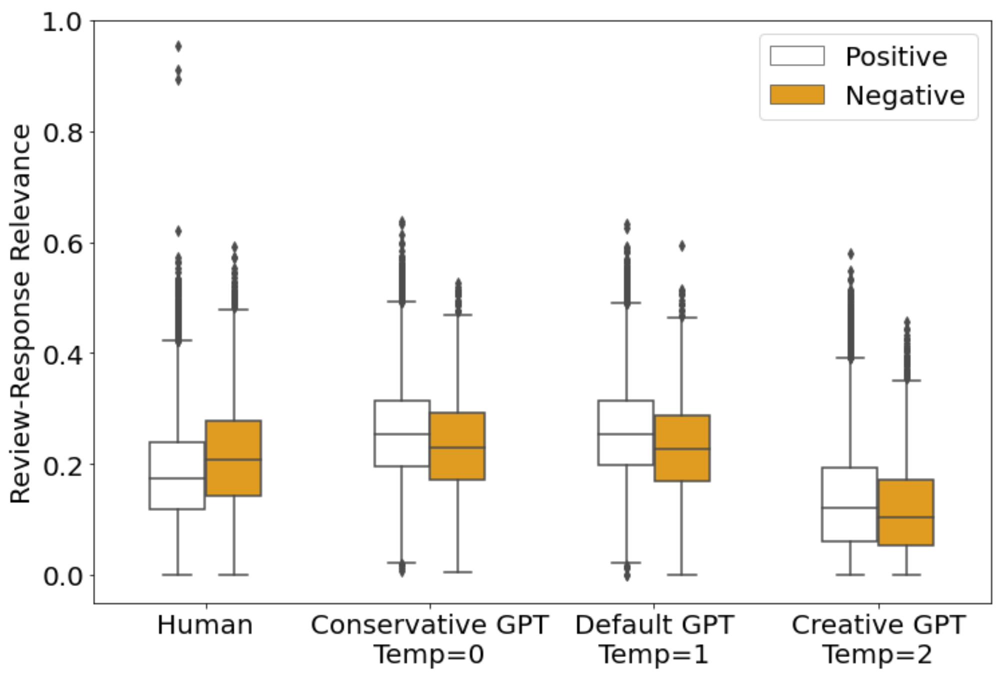

Figure 11. Degree of review-response relevance across different sources

Zhang H, Xiang Z, Zach FJ. Generative AI vs. humans in online hotel review management: A Task-Technology Fit perspective. *Tourism Management*.
DOI: 10.1016/j.tourman.2025.105187

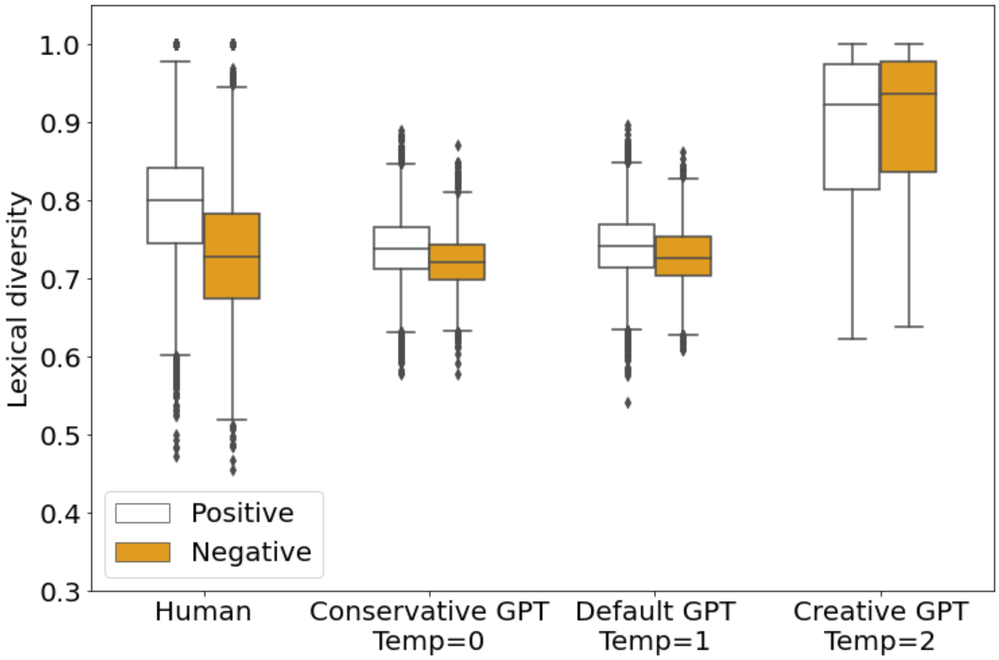

Figure 12. Degree of lexical diversity across different sources

To examine the relationship between human-AI similarity and review helpfulness under varying temperature settings, we estimated a group of regression models. They include a baseline model with only control variables and subsequent models that incorporate human-AI similarity and its interaction with review polarity. To mitigate the bias caused by the varying scales of variables, we standardized human-AI similarity and log transformed variables such as review title length and review content length. The results are displayed in Table 4. As shown in Model 1, our study concludes with results that are similar to previous studies, concerning the influence of several key factors such as review valence, review content length, review photo, response timeliness, and respondent role (K. L. Xie et al., 2017; Zhang et al., 2024).

In general, the default mode of GPT results in a negative outcome. When human responses resemble AI content with an increase of one standard deviation in human-AI similarity, their perceived helpfulness is likely to drop by 58% ($b = -0.58$, $p < 0.001$). Therefore, H1 is not supported. It indicates that the default mode of GPT is not effective for responding to online reviews disregard review valence. This finding is inconsistent with Litvin and Tan (2024), who observed that readers preferred GPT-generated responses. A possible explanation is that their study used iterative prompts to guide GPT in reducing redundancy in content and style. They reported that this refined prompting strategy outperformed the simpler approach which is used in our study. Further research is needed to explore how prompting strategies influence the effectiveness of Generative AI.

The effectiveness of Generative AI varies with technological attributes and task features. From the perspective of model temperatures, resembling AI responses at lower temperatures (0 or 1) is associated with worse customer perceptions. However, a creative mode (Temp = 2) is more beneficial, reflected by a positive and significant coefficient of human-AI similarity in model 6 (b = 0.18, p < 0.001). This is reasonable because predictable or formulaic replies, as produced by lower-temperature settings, might impair the perceived genuineness of the communication which is valued in the tourism industry (Hughes, 1995). However, high temperatures lead to more flexibility and variability, which signals companies' commitment to customer interactions (Wei et al., 2013), thus enhancing customer perceptions and attitudes.

When examining the alignment between model temperature and review valence, H2a (negative reviews) is not supported. The coefficients of human-AI similarity are -0.06, 0.01, and 0.03 under low, medium, and high temperatures, respectively. Higher temperature settings are more likely to amplify the positive effects of human-AI similarity. It indicates that the benefits of high temperatures, which enhance inter-review variation, outweigh the advantages of low temperatures. This finding aligns with previous studies, emphasizing the importance of high variation among responses, which has been recognized as a signal of a company's commitment to service recovery and customer communication (Wei et al., 2013; Zhang et al., 2020). Although high temperatures outperform the lower ones, the benefits of resembling AI remain stable across different temperature settings (Figures 13 to 15). This might be because lower temperatures are also featured with outputs that are effective for addressing complaints. As justified by previous studies, they can make the responses more relevant to the original review and concentrate on some specific points. These direct and thorough explanations improve the effectiveness of communication (Zhang et al., 2024; Zhang et al., 2020). Notably, the positive effect of the medium temperature in addressing negative reviews aligns with the findings of Tan et al. (2025), who used the same default GPT setting to generate responses to online complaints and found that customers showed a slight preference for AI-generated replies over those written by humans. This consistency reinforces the robustness of our findings, as they are backed by direct customer feedback from experimental studies.

In contrast, the empirical results support H2b. For positive reviews, AI models with low and medium temperatures (0 or 1) tend to yield adverse outcomes (Figures 13 and 14), while Figure 15 shows benefits of the high temperature setting (b = 0.22, p < 0.01). It indicates that the positive effect of human-AI similarity is stronger under a higher temperature. This finding is explainable because more conservative models tend to produce responses that are highly relevant to the original reviews, which could negatively impact customer perceptions. As is suggested by Deng and Ravichandran (2023), responses showing a high similarity to the original positive reviews can trigger the impression of overselling and provoke customer psychological reactance.

To test the validity of the findings, we conducted several robustness checks. First, we tested an alternative Generative AI model, Llama, which shares a similar transformer-model structure with GPT and shows strong capability of text generation, ensuring comparability to the GPT model (Touvron et al., 2023). The results using Llama are displayed in Appendix C. To address the imbalance between positive and negative reviews, we under-sampled positive reviews to construct a more balanced dataset. The results of this adjustment are shown in Appendix D. Additionally, given the high proportion of zeros in the dependent variable, we performed zero-inflated negative binomial regression (Zhang et al., 2024). The results of these

Zhang H, Xiang Z, Zach FJ. Generative AI vs. humans in online hotel review management: A Task-Technology Fit perspective. *Tourism Management*.
DOI: 10.1016/j.tourman.2025.105187

additional analyses can be found in Appendix E and F. Overall, the outcomes of these robustness checks demonstrate that our findings are consistent across different Generative AI systems, data sampling methods, and statistical models.

Zhang H, Xiang Z, Zach FJ. Generative AI vs. humans in online hotel review management: A Task-Technology Fit perspective. *Tourism Management*.
DOI: 10.1016/j.tourman.2025.105187

Table 4. Regression Results

	DV: Helpfulness						
	Baseline	**Conservative GPT** Temp = 0		**Default GPT** Temp = 1		**Creative GPT** Temp = 2	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Human-AI similarity		-0.61***	-0.71***	-0.58***	-0.69***	0.18***	0.22**
Human-AI similarity * Negative			0.65***		0.70***		-0.19*
Review Valence – Negative	0.55***	0.72***	0.85***	0.69***	0.80***	0.61***	0.56***
Review title length (log)	0.01	0.02	0.03	0.02	0.02	0.01	0.01
Review content length (log)	0.16**	0.17**	0.17**	0.23***	0.23***	0.21***	0.22***
Review readability (log)	0.04	0.11	0.12	0.12	0.14	0.04	0.04
Review photo	0.89***	1.02***	1.04***	1.02***	1.03***	0.84***	0.83***
Review recency (log)	-1.64***	-1.42***	-1.39***	-1.43***	-1.40***	-1.63***	-1.63***
Review timeliness (log)	-0.13	-0.16*	-0.17*	-0.15*	-0.16*	-0.04*	-0.04
Reviewer expertise (log)	0.15***	0.16***	0.17***	0.15***	0.16***	0.15***	0.15***
Response length (log)	0.41***	0.20**	0.17*	0.19*	0.17*	0.41***	0.41***
Response readability (log)	-0.07	009	0.06	0.10	0.07	-0.08	-0.09
Response sentiment	-0.64***	-0.18	-0.05	-0.18	-0.07	-0.67***	-0.67***
Response timeliness (log)	-0.26***	-0.30***	-0.31***	-0.30***	-0.30***	-0.27***	-0.27***
Respondent level – Administrative	0.63***	0.34***	0.33***	0.36***	0.35***	0.61***	0.61***
Respondent level – Executive	0.60**	0.58**	0.56*	0.54*	0.52*	0.61**	0.60**
Hotel review rating (log)	0.98	-1.15*	-1.04*	-0.98*	-0.88	0.87	0.92
Hotel review volume (log)	0.33***	0.14***	0.11**	0.16***	0.13***	0.33***	0.33***
Hotel class	-0.00	0.05	0.09	0.05	0.08	-0.02	-0.01
Hotel ranking (log)	-0.02	-0.15***	-0.16***	-0.13***	-0.14***	-0.03	-0.03
Intercept	-2.72**	1.65	1.51	1.05	0.89	-2.70**	-2.82**
Pseudo R^2	0.1499	0.173	0.177	0.172	0.176	0.153	0.154
Observations	32,129	32,129	32,129	32,129	32,129	32,129	32,129

Note: * p<0.05, ** p<0.01, *** p<0.001

Zhang H, Xiang Z, Zach FJ. Generative AI vs. humans in online hotel review management: A Task-Technology Fit perspective. *Tourism Management*.
DOI: 10.1016/j.tourman.2025.105187

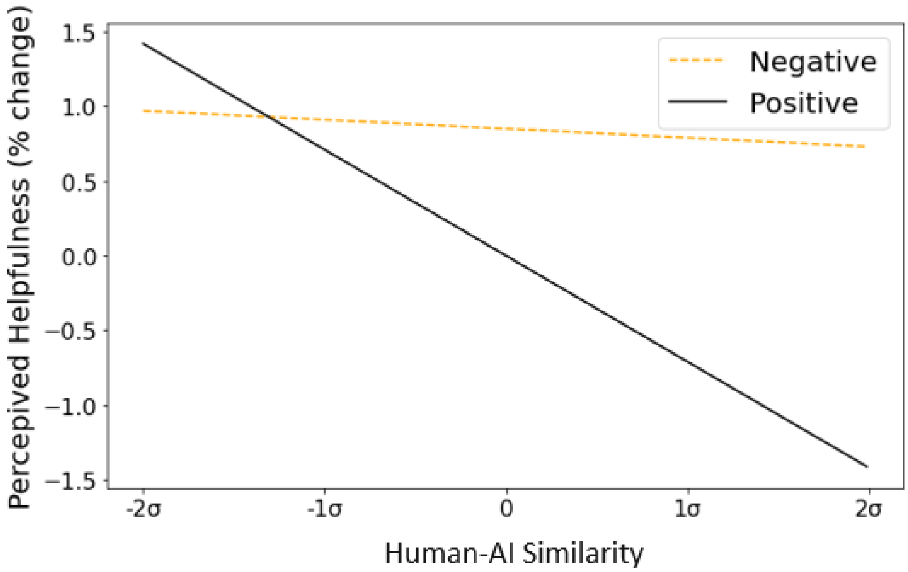

Figure 13. Estimated impact of Human-AI similarity: Conservative GPT (Temp = 0)

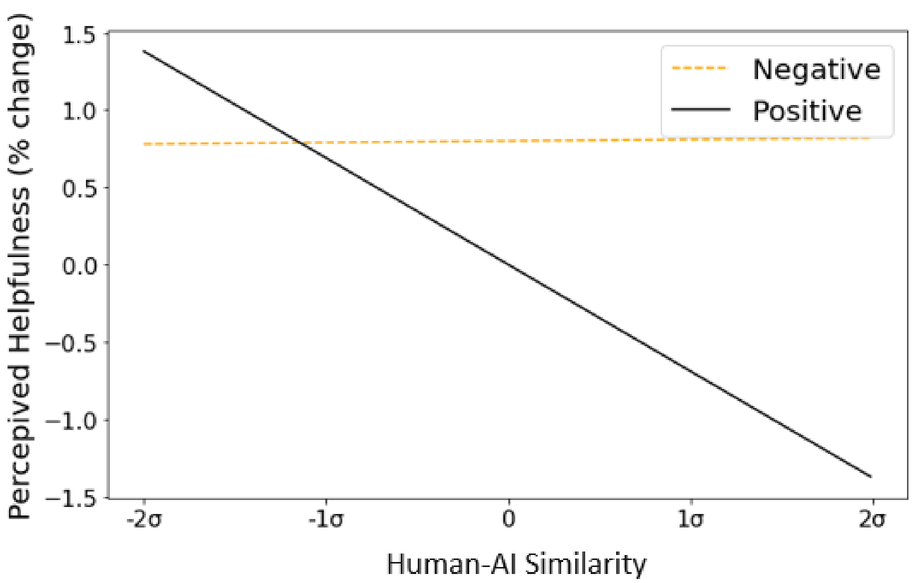

Figure 14. Estimated impact of Human-AI similarity: Default GPT (Temp = 1)

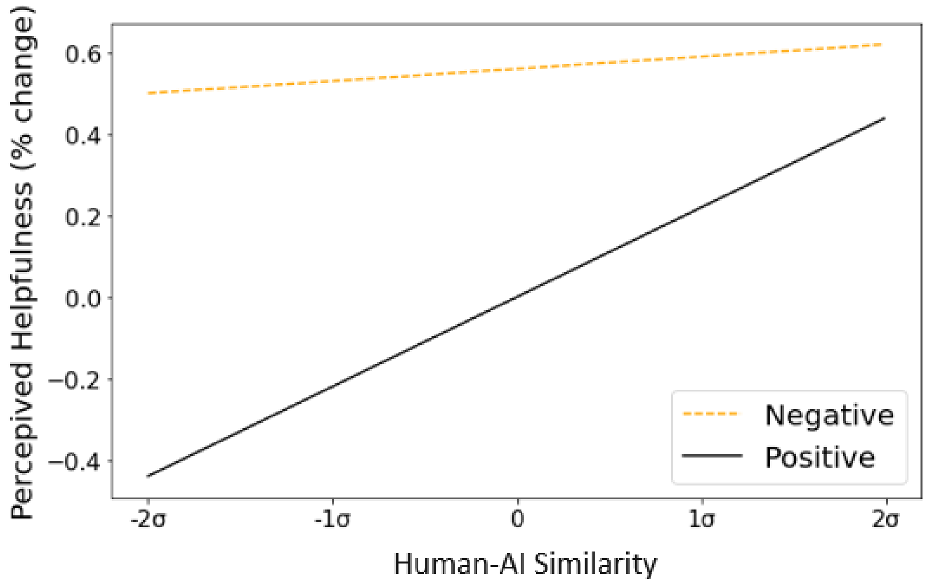

Figure 15. Estimated impact of Human-AI similarity: Creative GPT (Temp=2)

6. Implications and Limitations

This study examines the effectiveness of Generative AI in responding to online reviews, grounded in Task-Technology-Fit theory. The findings reveal that following the default mode of GPT will lead to lower perceived helpfulness, but this outcome varies with model temperature and review valence. High temperatures are more effective in addressing positive reviews while low and moderate temperatures are more suitable for negative reviews.

6.1 Theoretical Contributions and Implications

This research offers theoretical contributions and implications. First, it enriches tourism literature on Generative AI by introducing the lens of TTF based on big data analytics. This approach complements the preliminary research that focuses on conceptual discussions or empirical investigations without considering the dynamism of Generative AI (Gursoy et al., 2023; Tan et al., 2025). This study diverges from existing literature by focusing on Generative AI's adaptability and intelligence in responding to varied user inputs, specifically online reviews, at a large scale. This approach provides more reliable and generalizable findings by capturing the complexity and variability of this emerging technology, The findings, robust across different Generative AI models, also provide a reliable foundation for subsequent research in this field.

Second, by considering both task and technological features, the findings bring nuanced insights into Generative AI applications. The negative outcome under the default temperature setting regardless of review valence signifies the importance of adapting AI technique to specific review context. Higher temperature settings yield more favorable outcomes, especially for positive reviews, implying that variability in customer interactions is preferred in the service industry rather than repetitiveness (Liu et al., 2021). However, the risks associated with high temperatures, such as hallucination, are not covered in this study (Christensen et al., 2024). Future research could further explore this issue. By showing the outcomes of Generative AI adoption that are contingent on both task characteristics and technology features, this research can provide a foundation for subsequent research on Generative AI's adoption to broader customer-centered tasks beyond hospitality and tourism businesses. It emphasizes that Generative AI can be optimized by strategically configurating its features to align with specific task requirements.

Additionally, this study enriches our understanding of TTF theory by offering explanations for the alignment between task and technology through three content features: variation across responses, review-response relevance, and lexical diversity. Different tasks (i.e., review valence) demonstrate heterogeneous demands for each content feature, which can be altered through technological settings (i.e., temperature) (Bhavya et al., 2022; Xu & Zhao, 2022). These explanations substantiated by empirical evidence enhance the validity of the TTF theory.

Considering task, technology, and the mechanisms linking them, there are three potential directions for further extending this research. Future research could explore the effectiveness of Generative AI in more diverse tasks, such as social media content marketing and product description. This would allow for a comprehensive assessment of Generative AI’s adaptability to different content creation needs. From the technology perspective, subsequent studies are suggested to incorporate diverse technological settings of Generative AI models, such as prompt strategies and interaction modalities (e.g., voice versus text) (Xiaoyan et al., 2023). In terms of the alignment between task and technology, it is crucial to explore how different business tasks require specific content features, such as tone, emotion, credibility, and readability (Teixeira et al., 2012), that could be influenced by technological settings. Pursuing these research directions will help enrich the scope of TTF theory and advance our understanding of Generative AI’s capabilities in business applications.

6.2 Methodological and Practical Contributions

Methodologically, this study introduces a novel approach to explore the effectiveness of Generative AI by identifying whether human responses resembling AI patterns have received better outcomes. This simulated approach offers a way to test the potential outcomes of this technology without real-world adoption. In addition, the empirical tests based on big-data analytics complement traditional research methods, providing a more comprehensive understanding of how Generative AI performs across diverse hotel scenarios.

The practical implications of this study are manifold, offering valuable insights for both industry practitioners and policymakers. Overall, the results showcase the potential of applying Generative AI models like GPT and Llama to business operations. Generative AI could offer a solution to the persistent labor shortage in the tourism industry by automating repetitive and time-consuming tasks, allowing human staff to focus on more complex and personalized customer interactions.

For business operators within or beyond the tourism industry, this study provides valuable insights. The unwanted outcome associated with the default mode of GPT, which is widely used by individual workers now, underscores the necessity of developing internal guidelines Generative AI usage in online customer engagement at the workplace. Instead of relying on a one-size-fits-all approach, content creators should adapt their strategies. For example, the default mode of GPT can be used to address customer complaints. However, when replying to positive feedback, it is suggested to avoid using the default mode. Instead, staff can manually modify AI responses from the default mode to mimic the patterns of high temperatures, such as enhancing variation across responses and reducing the relevance to original texts. Alternatively, firms can purchase more professional Generative AI solutions. In addition, the complexity of aligning tasks and technological features highlights the need for workforce training. Employees should be equipped with knowledge of both business operations and AI capabilities to manage Generative AI effectively and unlock its full potential.

For AI system developers and solution providers, this research offers insight into creating more adaptable domain-specific Generative AI tools. The unsatisfactory outcomes of the default version of Generative AI tools, especially when addressing the dominant positive online reviews, underscore the need for more adaptable systems. By allowing users to adjust model parameters like temperature, these systems can better align with the demands of varying business tasks. Moreover, well-organized educational materials are needed to help their business users understand the adjustable parameters and how the adjustment can lead to different styles of AI-generated content. In addition, the difference in optimal settings across fields also highlights the need for domain-specific AI systems, such as fine-tuning existing AI systems with domain-specific datasets or training vertical domain large language models directly.

For policymakers, this research emphasizes the importance of regulatory frameworks that support the adaptable use of AI. Given the varying effectiveness of AI systems, policymakers should advocate for the development of industry-specific standards and best practices. They also need to ensure businesses have access to proper resources and training to facilitate effective and responsible Generative AI adoption. This includes policies that support AI training, performance monitoring, and continuous improvement to ensure the technology meets industry-specific needs.

6.3 Limitations

This study has several limitations. First, it only uses data from the Houston market, but results might differ in other regions due to varying cultural and business contexts. Second, the main relationships are tested on historical data without carrying out actual implementations. This approach offers valuable insights but may not fully capture the complexity of real-world adoption. Future studies should embrace field experiments to better understand the nuanced implementation of Generative AI and investigate their strategic outcomes. Third, this research does not involve human modifications to the GPT generated content, while in practice, businesses often edit AI responses before publishing to ensure alignment with their brand voice and customer service standards. The synergies between AI and humans and the tension between augmentation versus automation could further enrich this topic. Fourth, due to data availability constraints, our study relies on helpful votes as the primary dependent variable, which, while widely used in prior research, may not fully capture the effectiveness of managerial responses. Future research could explore alternative measures to provide a more comprehensive evaluation.

7. References

Bansal, G., Chamola, V., Hussain, A., Guizani, M., & Niyato, D. (2024). Transforming Conversations with AI—A Comprehensive Study of ChatGPT. *Cognitive Computation*, 1-24.

Bhavya, B., Xiong, J., & Zhai, C. (2022). Analogy generation by prompting large language models: A case study of instructgpt. *arXiv preprint arXiv:2210.04186*.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., & Askell, A. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877-1901.

Carlson, N. A. (2023). Differentiation in microenterprises. *Strategic Management Journal*, *44*(5), 1141-1167.

Choi, M., Choi, Y., Bangura, E., & Kim, D. (2024). ChatGPT or online review: Which is a better determinant of customers' trust in Airbnb listings and stay intention? *International Journal of Hospitality Management*, *122*, 103852.

Christensen, J., Hansen, J. M., & Wilson, P. (2024). Understanding the role and impact of Generative Artificial Intelligence (AI) hallucination within consumers' tourism decision-making processes. *Current Issues in Tourism*, 1-16.

Deng, C., & Ravichandran, T. (2023). Managerial Response to Online Positive Reviews: Helpful or Harmful? *Information Systems Research*.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Eriksen, A. V., Möller, S., & Ryg, J. (2024). Use of GPT-4 to diagnose complex clinical cases. In (Vol. 1, pp. AIp2300031): Massachusetts Medical Society.

Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS quarterly*, 213-236.

Gursoy, D., Li, Y., & Song, H. (2023). ChatGPT and the hospitality and tourism industry: an overview of current trends and future research directions. *Journal of Hospitality Marketing & Management*, *32*(5), 579-592.

Heya, T. A., Ineza, Y., Arefin, S. E., Uzor, G., & Serwadda, A. (2024). Stable or Shaky? The Semantics of ChatGPT's Behavior Under Repeated Queries. 2024 IEEE 18th International Conference on Semantic Computing (ICSC),

Hou, W., & Ji, Z. (2024). Comparing large language models and human programmers for generating programming code. *Advanced Science*, 2412279.

Hu, T., & Zhou, X.-H. (2024). Unveiling LLM Evaluation Focused on Metrics: Challenges and Solutions. *arXiv preprint arXiv:2404.09135*.

Huang, J., Gu, S. S., Hou, L., Wu, Y., Wang, X., Yu, H., & Han, J. (2022). Large language models can self-improve. *arXiv preprint arXiv:2210.11610*.

Huang, M.-H., & Rust, R. T. (2024). The caring machine: Feeling AI for customer care. *Journal of Marketing*, 00222429231224748.

Hughes, G. (1995). Authenticity in tourism. *Annals of Tourism Research*, *22*(4), 781-803.

Itani, O. S., Krush, M. T., Agnihotri, R., & Trainor, K. J. (2020). Social media and customer relationship management technologies: Influencing buyer-seller information exchanges. *Industrial Marketing Management*, *90*, 264-275.

Kim, J. H., Kim, J., Park, J., Kim, C., Jhang, J., & King, B. (2023). When ChatGPT gives incorrect answers: the impact of inaccurate information by generative AI on tourism decision-making. *Journal of Travel Research*, 00472875231212996.

Koneru, S., Wu, J., & Rajtmajer, S. (2023). Can Large Language Models Discern Evidence for Scientific Hypotheses? Case Studies in the Social Sciences. *arXiv preprint arXiv:2309.06578*.

Kwok, L. (2022). Labor shortage: a critical reflection and a call for industry-academia collaboration. *International Journal of Contemporary Hospitality Management*, *34*(11), 3929-3943.

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.

Li, C., Cui, G., & Peng, L. (2017). The signaling effect of management response in engaging customers: A study of the hotel industry. *Tourism Management*, *62*, 42-53.

Li, Y., & Lee, S. O. (2024). Navigating the generative AI travel landscape: the influence of ChatGPT on the evolution from new users to loyal adopters. *International Journal of Contemporary Hospitality Management*.

Liang, S., Zhang, X., Li, C., Li, H., & Yu, X. (2021). Tit for tat: understanding the responding behavior of property hosts on peer-to-peer rental platforms. *International Journal of Contemporary Hospitality Management*, *33*(3), 1105-1126.

Lin, H.-C., Han, X., Lyu, T., Ho, W.-H., Xu, Y., Hsieh, T.-C., Zhu, L., & Zhang, L. (2020). Task-technology fit analysis of social media use for marketing in the tourism and hospitality industry: a systematic literature review. *International Journal of Contemporary Hospitality Management*, *32*(8), 2677-2715.

Litvin, S. W., & Tan, K. P.-S. (2024). ChatGPT: It's Here, Whether We Want It or Not! *Cornell Hospitality Quarterly*, 19389655241229615.

Liu, S., Wang, N., Gao, B., & Gallivan, M. (2021). To be similar or to be different? The effect of hotel managers' rote response on subsequent reviews. *Tourism Management*, *86*, 104346.

Liu, X., Zheng, Y., Du, Z., Ding, M., Qian, Y., Yang, Z., & Tang, J. (2023). GPT understands, too. *AI Open*.

Lohr, S. (2024). *Generative A.I.'s Biggest Impact Will Be in Banking and Tech, Report Says*. https://www.nytimes.com/2024/02/01/business/ai-impact-jobs.html

Lopes, A. I., Dens, N., De Pelsmacker, P., & Malthouse, E. C. (2023). Managerial response strategies to eWOM: A framework and research agenda for webcare. *Tourism Management*, *98*, 104739.

Lu, L., Zhang, P., & Zhang, T. C. (2021). Leveraging "human-likeness" of robotic service at restaurants. *International Journal of Hospitality Management*, *94*, 102823.

Lucy, L., & Bamman, D. (2021). Gender and representation bias in GPT-3 generated stories. Proceedings of the third workshop on narrative understanding,

Ma, N. (2025). *Tripadvisor and Perplexity Partner to Redefine Travel Planning with AI Insights and Trusted Guidance*. https://www.hospitalitynet.org/news/4125323.html

Marr, B. (2023). *The Amazing Ways Expedia Is Using ChatGPT To Simplify Travel Arrangements*. https://www.forbes.com/sites/bernardmarr/2023/05/01/the-amazing-ways-expedia-is-using-chatgpt-to-simplify-travel-arrangements/?sh=74ced7663e7b

Mayer, H., Yee, L., Chui, M., & Roberts, R. (2025). *Superagency in the workplace: Empowering people to unlock AI's full potential*. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/superagency-in-the-workplace-empowering-people-to-unlock-ais-full-potential-at-work

McKinsey. (2023). *The economic potential of generative AI: The next productivity frontier*. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#/

Microsoft. (2024). *Microsoft and LinkedIn released 2024 Work Trend Index: three out of four people use AI at work*. https://news.microsoft.com/en-cee/2024/05/13/microsoft-and-linkedin-released-2024-work-trend-index-three-out-of-four-people-use-ai-at-work/

Min, H., Lim, Y., & Magnini, V. P. (2015). Factors affecting customer satisfaction in responses to negative online hotel reviews: The impact of empathy, paraphrasing, and speed. *Cornell Hospitality Quarterly*, *56*(2), 223-231.

Morosan, C. (2025). Rethinking information disclosure to GenAI in hotels: An extended parallel process model. *International Journal of Hospitality Management*, *124*, 103965.

Mukherjee, P., Hou, B., Lanfredi, R. B., & Summers, R. M. (2023). Feasibility of using the privacy-preserving large language model Vicuna for labeling radiology reports. *Radiology*, *309*(1), e231147.

Nicolau, J. L., Xiang, Z., & Wang, D. (2024). Daily online review sentiment and hotel performance. *International Journal of Contemporary Hospitality Management*, *36*(3), 790-811.

Paris, M. (2023). *ChatGPT Hits 100 Million Users, Google Invests In AI Bot And CatGPT Goes Viral*. https://www.forbes.com/sites/martineparis/2023/02/03/chatgpt-hits-100-million-microsoft-unleashes-ai-bots-and-catgpt-goes-viral/

Przegalinska, A., Triantoro, T., Kovbasiuk, A., Ciechanowski, L., Freeman, R. B., & Sowa, K. (2025). Collaborative AI in the workplace: Enhancing organizational performance through resource-based and task-technology fit perspectives. *International Journal of Information Management*, *81*, 102853.

Ravichandran, T., & Deng, C. (2023). Effects of managerial response to negative reviews on future review valence and complaints. *Information Systems Research*, *34*(1), 319-341.

Reisenbichler, M., Reutterer, T., Schweidel, D. A., & Dan, D. (2022). Frontiers: Supporting content marketing with natural language generation. *Marketing Science*, *41*(3), 441-452.

Rivera, M., Croes, R., & Zhong, Y. (2016). Developing mobile services: A look at first-time and repeat visitors in a small island destination. *International Journal of Contemporary Hospitality Management*, *28*(12), 2721-2747.

Russell, S. J., & Norvig, P. (2010). *Artificial intelligence a modern approach*. London.

Shin, H., Perdue, R. R., & Pandelaere, M. (2020). Managing customer reviews for value co-creation: An empowerment theory perspective. *Journal of Travel Research*, *59*(5), 792-810.

Shin, H. H., & Jeong, M. (2022). Redefining luxury service with technology implementation: the impact of technology on guest satisfaction and loyalty in a luxury hotel. *International Journal of Contemporary Hospitality Management*, *34*(4), 1491-1514.

Shin, S., Du, Q., Ma, Y., Fan, W., & Xiang, Z. (2021). Moderating effects of rating on text and helpfulness in online hotel reviews: an analytical approach. *Journal of Hospitality Marketing & Management*, *30*(2), 159-177.

Shin, S., Kim, J., Lee, E., Yhee, Y., & Koo, C. (2023). ChatGPT for trip planning: the effect of narrowing down options. *Journal of Travel Research*, 00472875231214196.

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.

Teixeira, T., Wedel, M., & Pieters, R. (2012). Emotion-induced engagement in internet video advertisements. *Journal of Marketing Research*, *49*(2), 144-159.

Touch, G. (2025). *Next Generation Reputation, Messaging & Guest Engagement Platform*. https://www.guesttouch.com/

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., & Azhar, F. (2023). Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine Learning research*, *9*(11).

Van Dis, E. A., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: five priorities for research. *nature*, *614*(7947), 224-226.

Wang, B., & Jia, T. (2023). Seize the favorable impression: how hosts should manage positive online reviews. *International Journal of Contemporary Hospitality Management*.

Wei, W., Miao, L., & Huang, Z. J. (2013). Customer engagement behaviors and hotel responses. *International Journal of Hospitality Management*, *33*, 316-330.

Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management*, *31*(2), 179-188.

Xiaoyan, Z., Yang, D., Min, Y., Lingzhi, W., Rui, Z., Hong, C., Wai, L., Ying, S., & Ruifeng, X. (2023). A comprehensive survey on deep learning for relation extraction: Recent advances and new frontiers. *arXiv preprint arXiv:2306.02051*.

Xie, K., Kwok, L., & Wang, W. (2017). Monetizing managerial responses on TripAdvisor: Performance implications across hotel classes. *Cornell Hospitality Quarterly*, *58*(3), 240-252.

Xie, K. L., So, K. K. F., & Wang, W. (2017). Joint effects of management responses and online reviews on hotel financial performance: A data-analytics approach. *International Journal of Hospitality Management*, *62*, 101-110.

Xu, S., Zhang, X., Kim, R., & Su, M. (2024). Anthropomorphic last-mile robots and consumer intention: An empirical test under a theoretical framework. *Journal of Retailing and Consumer Services*, *81*, 104028.

Xu, X., & Zhao, Y. (2022). Examining the influence of linguistic characteristics of online managerial response on return customers' change in satisfaction with hotels. *International Journal of Hospitality Management*, *102*, 103146.

Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H., Zhong, S., Yin, B., & Hu, X. (2023). Harnessing the power of llms in practice: A survey on chatgpt and beyond. *ACM Transactions on Knowledge Discovery from Data*.

Zhang, X., La, L., Huang, G. I., & Xie, H. (2024). Strategies and conditions for crafting managerial responses to online reviews. *Tourism Management*, *103*, 104911.

Zhang, X., Qiao, S., Yang, Y., & Zhang, Z. (2020). Exploring the impact of personalized management responses on tourists' satisfaction: A topic matching perspective. *Tourism Management*, *76*, 103953.

Zhang, Z., Li, H., Meng, F., & Li, Y. (2019). The effect of management response similarity on online hotel booking.

Zhou, S., Yan, Q., Yan, M., & Shen, C. (2020). Tourists' emotional changes and eWOM behavior on social media and integrated tourism websites. *International Journal of Tourism Research*, *22*(3), 336-350.

Zhou, T., Wang, Y., Yan, L., & Tan, Y. (2023). Spoiled for choice? Personalized recommendation for healthcare decisions: A multiarmed bandit approach. *Information Systems Research*, *34*(4), 1493-1512.

Ziegler, D. M., Stiennon, N., Wu, J., Brown, T. B., Radford, A., Amodei, D., Christiano, P., & Irving, G. (2019). Fine-tuning language models from human preferences. *arXiv preprint arXiv:1909.08593*.