Tourist Satisfaction Index Development - A Large Language Model Approach

Haiyan Song, **Changhua He** & **Curely Liu**

Research Centre for Digital Transformation of Tourism

School of Hotel and Tourism Management

The Hong Kong Polytechnic University

Abstract: This study presents a novel method for compiling the Tourist Satisfaction Index (TSI) using Large Language Models (LLMs) to analyze online review data. Building on prior research, we develop a theoretical framework that integrates tourism service quality, tourists' cognitive-affective responses, and satisfaction. LLMs are employed to quantify these relationships by analyzing multilingual review data from social media and online travel agencies (OTAs), offering an alternative to traditional questionnaire-based TSI assessments. The method is empirically validated using hotel review data from Hong Kong, demonstrating its effectiveness in generating granular, sector-specific insights for service improvement.

Keywords: Tourist Satisfaction Index, SERVQUAL model, cognitive-affective response, online reviews, large language models (LLMs)

1. Introduction

Tourist satisfaction is critical for destination competitiveness, yet traditional questionnaire-based methods suffer from preset attributes and data silos. Online reviews provide rich, real-time data to overcome these limitations. This study leverages LLMs to analyze reviews, capturing both cognitive and affective dimensions of satisfaction, and proposes a scalable framework for TSI compilation.

2. Theoretical framework

The framework (Figure 1) integrates SERVQUAL dimensions (Tangibles, Reliability, Responsiveness, Assurance, Empathy) and price to explain how service quality influences cognitive-affective responses and satisfaction. Online reviews act as signals of tourist perceptions, with linguistic features analyzed via LLMs to quantify satisfaction.

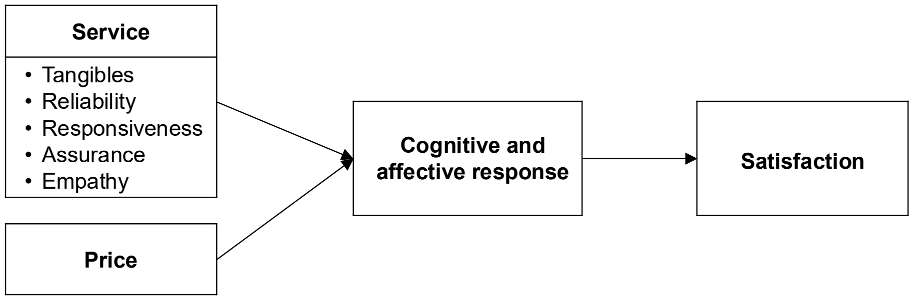

Figure 1. The TSI assessment framework.

Figure 1 suggests service and price are the key factors that affect tourists' cognitive and affective responses to the service quality (Song et al., 2012; Shokouhyar et al., 2020), which is further influence tourists' satisfaction (Song et al., 2011, 2012). As per the SERVQUAL model (Parasuraman et al., 1988), the tourism service quality can be categorized into five dimensions: Tangibles, Reliability, Responsiveness, Assurance, and Empathy, as shown in Table 1. In addition to service quality, the price of tourism products/services also plays an important role in determining tourist satisfaction, as it reflects tourists' assessed value of the product/service they receive.(Song et al., 2011, 2012).

The experience-sharing behavior of tourists is typically manifested through the review data, which reflects their cognitive and affective perceptions of tourism products/services. Firstly, Personal Construct Theory (PCT) highlights an individual's inclination to share pertinent and valid information with others (Kelly, 1963; Kim et al., 2022), encompassing their attitudes and opinions regarding the travel experience (Kawaf & Tagg, 2017). This sharing behavior fulfills tourists' affective needs. Secondly, following the Uses and Gratifications Theory (UGT), tourists perceive social media and OTA platforms as ideal sharing channels (Leung, 2013). Online information sharing (e.g., online reviews) stimulates tourists' gratification in sharing behavior through a combination of social affection, cognitive needs, the need to express negative emotions and feelings, recognition, and entertainment (Leung, 2013; Kim et al., 2022). For these reasons, online reviews are considered to be more reflective than the pre-designed surveys due to their capacity to harness extensive and real-experience-related data (Liu et al., 2017; Shin & Nicolau, 2022), and offer valuable multi-dimensional insights into tourists' perceptions, expectations, emotions, and attitudes (Taecharungroj & Mathayomchan, 2019; Jia, 2020; Li et al., 2020; Wang & Li, 2024).

Table 1 Dimensions of service and topics.

Dimensions		Topics
Service quality	Tangibles	Facilities and Equipment Environment and Atmosphere Location
	Reliability	Accuracy of Service Consistency and Persistence of Services
	Responsiveness	Timeliness of Response Employees' Willingness to help
	Assurance	Professionalism of Employees Perceived Safety Manner of Employees
	Empathy	Personalized Service Attention to Customers
Price	Price	Assessed Value

From the perspective of signaling theory, tourists must rely on the content and mode of reviews for effective information transmission (Spence, 2002; Connelly et al., 2011; Kim et al., 2022). Hence, the semantic and linguistic features exhibited in these reviews can serve as crucial signals to reflect tourists' perceptions regarding the quality and pricing of tourism products/services (Geetha et al., 2017; Carvalho & Ivanov, 2024; Zhang et al., 2024a). Although many different models have been developed in the natural language processing field, LLM stands out as the most advanced technique due to its exceptional capability to handle intricate and multilingual textual data (Vaswani et al., 2017; Zhang et al., 2024b). Through pre-training on extensive textual data, LLMs have acquired prior knowledge of lexical information, semantic comprehension, and language logic in natural language (Carvalho & Ivanov, 2024; Zhang et al., 2024b). In the context of review data, LLMs can extract satisfaction information by leveraging semantic and linguistic features. This study therefore uses LLMs to capture information on tourists' cognitive-affective responses (Dai et al., 2015), quantify satisfaction levels (Carvalho & Ivanov, 2024; Zhang et al., 2024a), and ultimately compile TSIs.

3. Methodology

Based on the review data, the proposed four-stage assessment system first aggregates TSIs at the topical level, then further aggregates these for the sectors, and finally, the destination TSIs can be calculated, as shown in Figure 2.

Stage 1, data collection and preprocessing. This stage employs web crawler technology to collect the online review data. Recognizing that tourists often cover multiple topics within a single review (as depicted in Figure 3), this study employed sentence segmentation techniques to ensure each sentence focuses only on one topic.

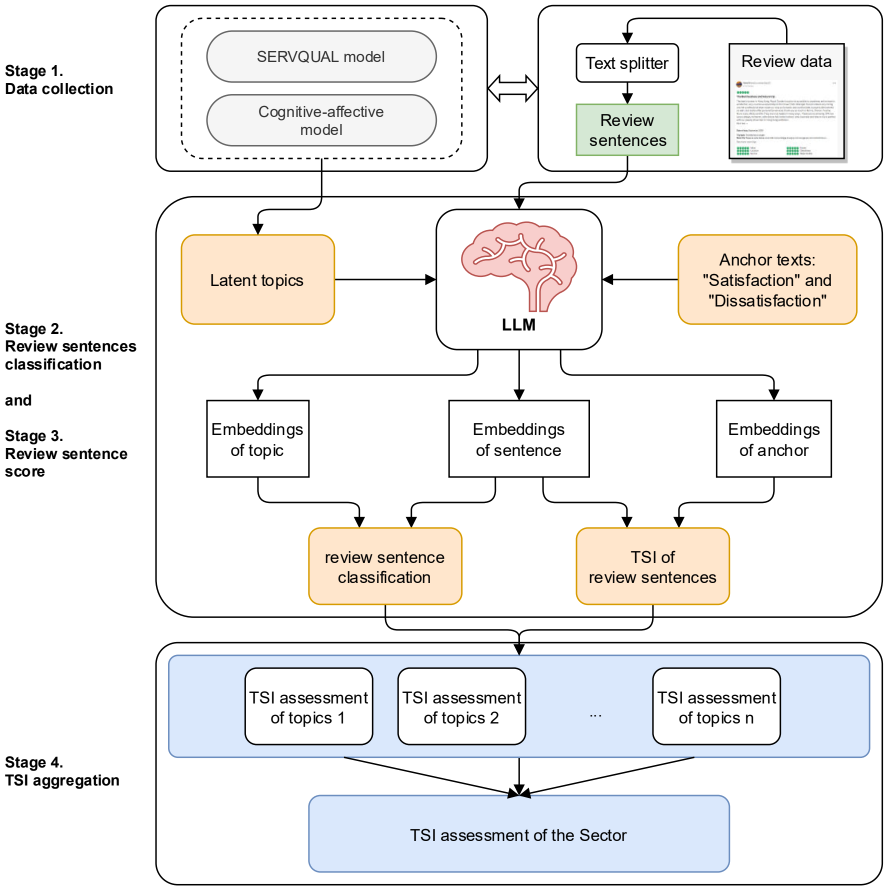

Figure 2. The bottom-up TSI assessment method.

In Stages 2 and 3, a pre-trained LLM, the General Text Embedding (GTE) model1 (Zhang et al., 2024b), was utilized as an embedded tool to extract semantic and linguistic information from the review text. Initially, the GTE model represents topics, reviews, and anchor words as dense feature vectors. In Stage 2, we employ cosine similarity between topic and review vectors to classify reviews into different topics. In Stage 3, we quantify the tourists' propensity score to satisfaction by measuring the similarity between the anchor and review vectors. It's worth noting that the inherent tolerance of tourists may lead them to assign a satisfactory rating, even if they mention dissatisfactory content in their review text, as illustrated in Figure 4. We adjust the propensity score using the optimal classification threshold determined by Kolmogorov-Smirnov (KS) (Yu & He, 2023). The adjusted propensity score of satisfaction is normalized within a range of 0-100 to represent actual satisfaction scores.

[1] More details are available at https://huggingface.co/Alibaba-NLP/gte-multilingual-base

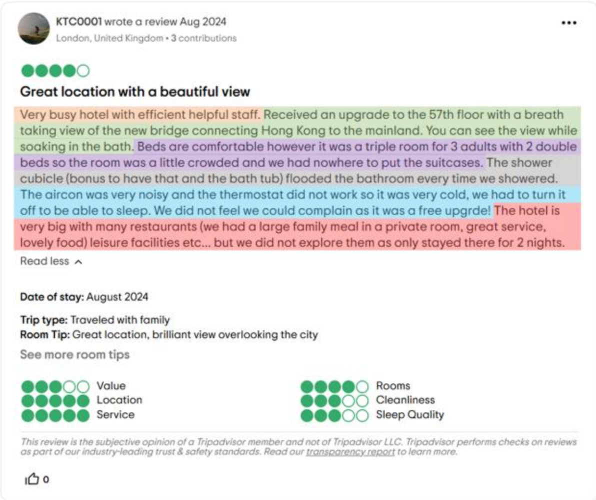

Figure 3. Screenshot of a sample review displaying multiple topics (TripAdvisor.com).

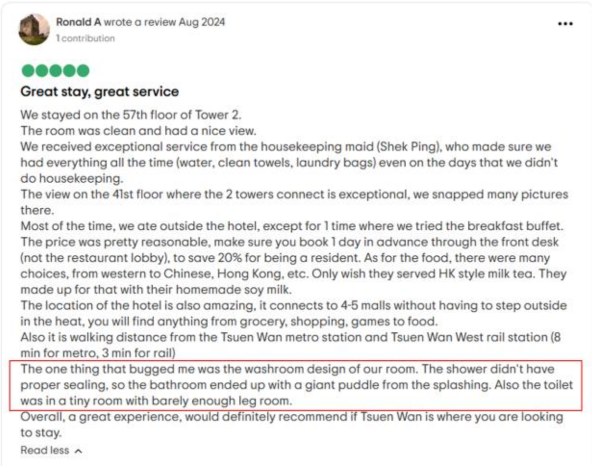

Figure 4. Screenshot of a sample review with unsatisfactory content but a 5-star rating (TripAdvisor.com).

Stage 4, TSI aggregation. We employ a bottom-up aggregation strategy in this stage to derive the final sectoral TSI. The implementation process involves two steps. First, the TSIs at the topic level were obtained by averaging the satisfaction scores of all review texts for each topic. Second, the sectoral TSI is calculated by aggregating the topic TSI scores, weighted by the number of reviews within each topic.

4. Empirical results

This study utilizes the hotel industry in Hong Kong as a case study to conduct the TSI assessment, with data sourced from TripAdvisor, a renowned review platform in the travel domain. At the time of this study, 1,502 hotels in Hong Kong were listed on TripAdvisor, out of which 1,003 hotels contained tourist review data. Employing web crawler technology, we collected a total of 407,922 review entries.

To evaluate the suitability of the proposed framework, we conducted a classification experiment using raw review data. In this experiment, reviews rated 1 and 2 stars were labeled as "unsatisfactory," while reviews rated 4 and 5 stars were labeled as "satisfactory." The classification experiment employs Area Under the Curve (AUC) as the evaluation metric, and the results are illustrated in Figure 5. As depicted in the Figure, GTE achieved the highest AUC value among all LLMs utilized, reaching 0.943. This can be attributed to GTE's extensive multilingual capability (supporting 70 languages), which aligns with the characteristics of multilingual review data. Additionally, this may also stem from two core components (i.e., RoPE and Unpadding), which enable GTE to adapt to intricate user-generated comment data effectively (Su et al., 2024). Interestingly, the optimal classification threshold for satisfaction and dissatisfaction is determined to be -0.083. This finding implies that tourists exhibit a certain level of tolerance by assigning a 5-star rating despite expressing dissatisfaction in their reviews. To mitigate potential overestimation, this paper evaluates each sentence independently, thereby addressing concerns regarding hoteliers' overoptimism.

Based on the assumption that the number of reviews reflects the attention paid by tourists to the topic, we computed the weights for each theme, as presented in Table 2. The result indicates that "Willingness of employees" and "Manner of employees" have the highest weights at 0.3323 and 0.4259, respectively. Notably, both themes highlight the interaction between tourists and hotels, suggesting that the topics involving greater engagement are more likely to leave an impression on tourists and consequently receive more attention in post-consumer reviews.

The results of the TSI assessment for Hong Kong's hotel industry are presented in Figure 6. As depicted in Figure 6, the TSI for Hong Kong's hotel industry exhibits a consistent upward trend over the years, indicating notable advancements in service quality within the hotel sector. However, it is noteworthy that post-2020, there appears to be a relatively slower growth in the TSI index, possibly attributed to the impact of the COVID-19.

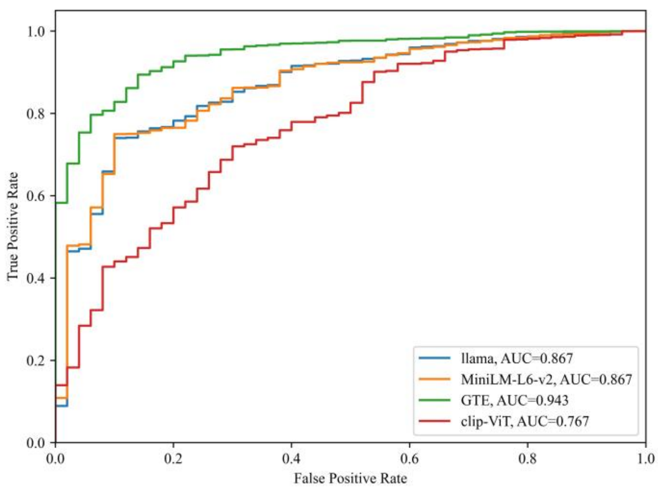

Figure 5. AUC index diagram of the proposed framework.

Table 2 Weights of topics.

Dimensions		Topics	Counts	Weights
Service quality	Tangibles (0.0615)	Facilities and Equipment	33628	0.0122
		Environment and Atmosphere	99549	0.0362
		Location	35949	0.0131
	Reliability (0.0420)	Accuracy of Service	6112	0.0022
		Consistency and Persistence of Services	109295	0.0398
	Responsiveness (0.3328)	Timeliness of Response	1386	0.0005
		Employees' Willingness to help	913351	0.3323
	Assurance (0.4313)	Professionalism of Employees	7311	0.0027
		Perceived Safety	7669	0.0028
		Manner of Employees	1170558	0.4259
	Empathy (0.0856)	Personalized Service	3442	0.0013
		Attention to Customers	231770	0.0843
Price	Price (0.0468)	Assessed Value	128549	0.0468

By focusing on the topics, it can be observed that these 13 topics can be categorized into three groups, as illustrated in Figure 7. The first group encompasses the topic "Willingness of employees", which exhibits a significantly higher satisfaction than the others and demonstrates a consistent upward trend. The second group comprises three topics, "Environment and Atmosphere", "Personalized Service", and "Manner of Employees", displaying a gradual improvement over time. However, there have been relatively significant fluctuations since 2020, with the topic of "Personalized Service" being more volatile. This could potentially be attributed to COVID-19, where hotels prefer standardized and secure services while reducing their focus on "Personalized Service". All other topics fall under the third category, exhibiting a steady overall trend. Notably, the topics "Perceived Safety" and "Professionalism of Employees" warrant attention: the satisfaction of the topic "Perceived Safety" is around 50%, indicating tourists' conservative attitude towards this theme without any explicit satisfaction. The topic "Professionalism of Employees" has experienced a significant decline post-2020, reflecting hotel practitioners' inability to meet tourists' service expectations.

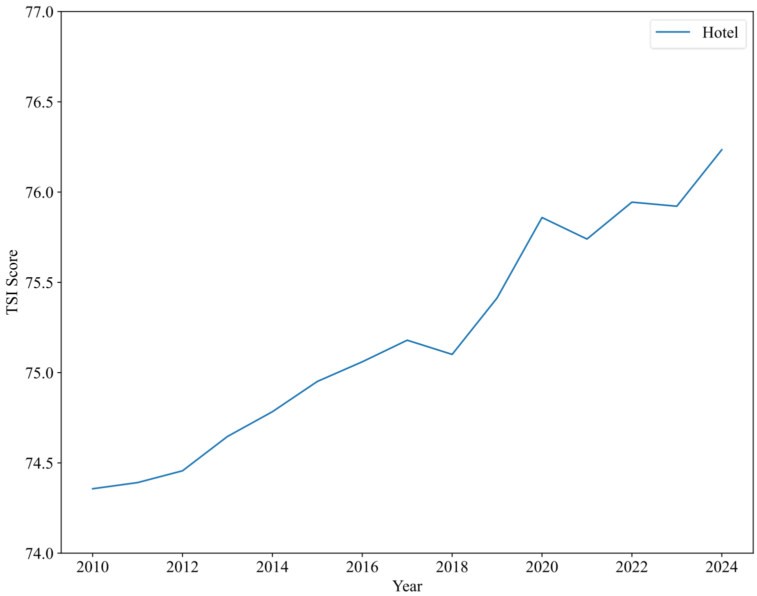

Figure 6. TSI assessment of the hotel industry in Hong Kong.

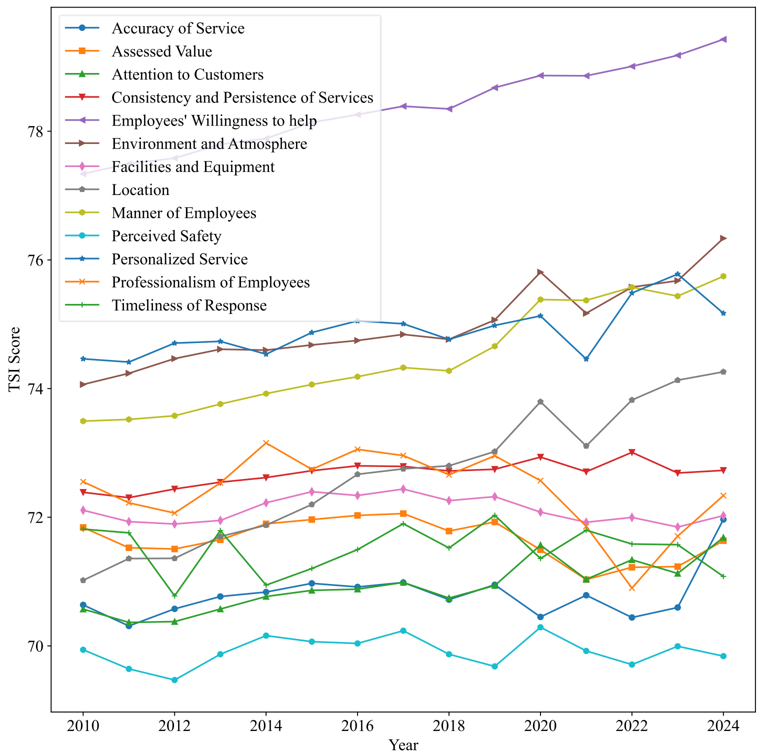

Figure 7. Topical-level TSI assessment.

The differences in satisfaction by topics within the context of trip types and languages are illustrated in Figure 8. The satisfactions are more heterogeneous among different language backgrounds than among trip types. Similar to Figure 7, the topic "Willingness to help of Employees" receives the highest level of satisfaction from most tourists, while the topic "Perceived Safety" garners the least satisfaction. Notably, distinct satisfaction patterns emerge among various types of tourists. Specifically, tourists from English-speaking countries (denoted as "en") exhibit higher expectations across topics "Perceived Safety", "Attention to Customers", and "Accuracy of

Service", resulting in lower levels of satisfaction. This could be attributed to more significant cultural disparities in English-speaking nations, leading to a relatively higher number of dissatisfactory aspects across multiple topics. Similarly, Chinese-speaking individuals (denoted as "zhCN") also express low levels of satisfaction with several topics such as "Timeliness of Response", "Perceived Safety", and "Accuracy of Service" due to similar reasons. Conversely, Japanese tourists (denoted as "ja") indicate dissatisfaction primarily with the topic "Accuracy of Service", and Korean tourists (denoted as "ko") particularly focus on the topic "Accuracy of Service".

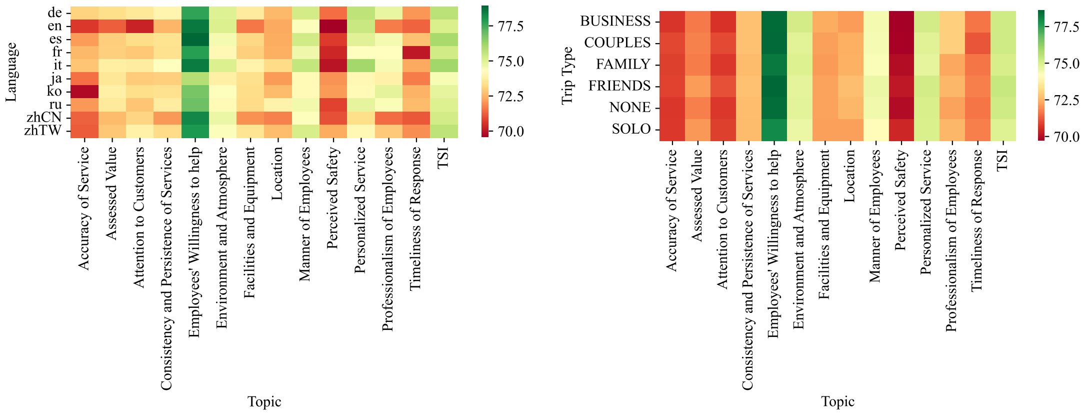

Figure 8. The impact of language and trip type on satisfaction.

5. Concluding remarks

This study proposes an alternative TSI assessment method incorporating both cognitive and affective responses based on review data and LLMs. By taking the hotel sector in Hong Kong as an example, this study demonstrates the method's suitability in sectoral TSI assessment. The results show that the hotel industry in Hong Kong has exhibited a consistent upward increase in TSI. Moreover, the results also reveal that tourists primarily emphasize two interactive topics: "Willingness to help of Employees" and "Manner of Employees".

Theoretically, this study proposes an innovative approach by integrating the review data and LLM with a cognitive-affective TSI assessment framework. One of the managerial implications of this study is that hotels can implement precise service improvement strategies based on the TSI assessment results for different topics. Moreover, the assessment framework can be applied in other tourism sectors such as attractions, restaurants, retail shops, transportation, and immigration services, which allows the compilation of TSIs at the topic, sectoral, and destination levels.

References

Bagozzi, R. P., Gopinath, M., & Nyer, P. U. (1999). The Role of Emotions in Marketing. *Journal of the Academy of Marketing Science*, *27*(2), 184–206. https://doi.org/10.1177/0092070399272005

Carvalho, I., & Ivanov, S. (2024). ChatGPT for tourism: applications, benefits and risks. *Tourism Review*, *79*(2), 290–303. https://doi.org/10.1108/TR-02-2023-0088

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling Theory: A Review and Assessment. *Journal of Management*, *37*(1), 39–67. https://doi.org/10.1177/0149206310388419

Dai, H., Luo, X. (Robert), Liao, Q., & Cao, M. (2015). Explaining consumer satisfaction of services: The role of innovativeness and emotion in an electronic mediated environment. *Decision Support Systems*, *70*, 97–106. https://doi.org/10.1016/j.dss.2014.12.003

Del Bosque, I. R., & Martín, H. S. (2008). Tourist satisfaction a cognitive-affective model. *Annals of Tourism Research*, *35*(2), 551–573. https://doi.org/10.1016/j.annals.2008.02.006

Frijda, N. H. (1988). The Laws of Emotion. *American Psychologist*, *43*(5), 349–358. Scopus. https://doi.org/10.1037/0003-066X.43.5.349

Geetha, M., Singha, P., & Sinha, S. (2017). Relationship between customer sentiment and online customer ratings for hotels - An empirical analysis. *Tourism Management*, *61*, 43–54. https://doi.org/10.1016/j.tourman.2016.12.022

González-Rodríguez, M. R., Díaz-Fernández, M. C., & Pacheco Gómez, C. (2020). Facial-expression recognition: An emergent approach to the measurement of tourist satisfaction through emotions. *Telematics and Informatics*, *51*, 101404. https://doi.org/10.1016/j.tele.2020.101404

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

Huang, S. (Sam), Weiler, B., & Assaker, G. (2015). Effects of Interpretive Guiding Outcomes on Tourist Satisfaction and Behavioral Intention. *Journal of Travel Research*, *54*(3), 344–358. https://doi.org/10.1177/0047287513517426

Jia, S. (Sixue). (2020). Motivation and satisfaction of Chinese and U.S. tourists in restaurants: A cross-cultural text mining of online reviews. *Tourism Management*, *78*, 104071. https://doi.org/10.1016/j.tourman.2019.104071

Kawaf, F., & Tagg, S. (2017). The construction of online shopping experience: A repertory grid approach. *Computers in Human Behavior*, *72*, 222–232. https://doi.org/10.1016/j.chb.2017.02.055

Kelly, G. (1963). *A theory of personality: The psychology of personal constructs*. WW Norton & Company.

Kim, J., Lee, M., Kwon, W., Park, H., & Back, K.-J. (2022). Why am I satisfied? See my reviews – Price and location matter in the restaurant industry. *International Journal of Hospitality Management*, *101*, 103111. https://doi.org/10.1016/j.ijhm.2021.103111

Leung, L. (2013). Generational differences in content generation in social media: The roles of the gratifications sought and of narcissism. *Computers in Human Behavior*, *29*(3), 997–1006. https://doi.org/10.1016/j.chb.2012.12.028

Li, H., Liu, Y., Tan, C.-W., & Hu, F. (2020). Comprehending customer satisfaction with hotels: Data analysis of consumer-generated reviews. *International Journal of Contemporary Hospitality Management*, *32*(5), 1713–1735.

Liu, Y., Teichert, T., Rossi, M., Li, H., & Hu, F. (2017). Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews. *Tourism Management*, *59*, 554–563. https://doi.org/10.1016/j.tourman.2016.08.012

Oliver, R. L. (1993). Cognitive, Affective, and Attribute Bases of the Satisfaction Response. *Journal of Consumer Research*, *20*(3), 418–430. https://doi.org/10.1086/209358

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perc. *Journal of Retailing*, *64*(1), 12.

Prayag, G., Hosany, S., Muskat, B., & Del Chiappa, G. (2017). Understanding the Relationships between Tourists' Emotional Experiences, Perceived Overall Image, Satisfaction, and Intention to Recommend. *Journal of Travel Research*, *56*(1), 41–54. https://doi.org/10.1177/0047287515620567

Rodríguez Molina, M. Á., Frías-Jamilena, D.-M., & Castañeda-García, J. A. (2013). The moderating role of past experience in the formation of a tourist destination's image and in tourists' behavioural intentions. *Current Issues in Tourism*, *16*(2), 107–127. https://doi.org/10.1080/13683500.2012.665045

Shin, S., & Nicolau, J. L. (2022). Identifying attributes of wineries that increase visitor satisfaction and dissatisfaction: Applying an aspect extraction approach to online reviews. *Tourism Management*, *91*, 104528. https://doi.org/10.1016/j.tourman.2022.104528

Shokouhyar, S., Shokoohyar, S., & Safari, S. (2020). Research on the influence of after-sales service quality factors on customer satisfaction. *Journal of Retailing and Consumer Services*, *56*, 102139. https://doi.org/10.1016/j.jretconser.2020.102139

Sirakaya, E., Petrick, J., & Choi, H.-S. (2004). The Role of Mood on Tourism Product Evaluations. *Annals of Tourism Research*, *31*(3), 517–539. https://doi.org/10.1016/j.annals.2004.01.009

Song, H., Li, G., van der Veen, R., & Chen, J. L. (2011). Assessing mainland Chinese tourists' satisfaction with Hong Kong using tourist satisfaction index. *International Journal of Tourism Research*, *13*(1), 82–96. https://doi.org/10.1002/jtr.801

Song, H., Van Der Veen, R., Li, G., & Chen, J. L. (2012). The Hong Kong tourist satisfaction index. *Annals of Tourism Research*, *39*(1), 459–479. https://doi.org/10.1016/j.annals.2011.06.001

Spence, M. (2002). Signaling in Retrospect and the Informational Structure of Markets. *American Economic Review*, *92*(3), 434–459. https://doi.org/10.1257/00028280260136200

Spreng, R. A., MacKenzie, S. B., & Olshavsky, R. W. (1996). A Reexamination of the

Determinants of Consumer Satisfaction. *Journal of Marketing*, *60*(3), 15–32. https://doi.org/10.1177/002224299606000302

Su, J., Ahmed, M., Lu, Y., Pan, S., Bo, W., & Liu, Y. (2024). RoFormer: Enhanced transformer with Rotary Position Embedding. *Neurocomputing*, *568*, 127063. https://doi.org/10.1016/j.neucom.2023.127063

Taecharungroj, V., & Mathayomchan, B. (2019). Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. *Tourism Management*, *75*, 550–568. https://doi.org/10.1016/j.tourman.2019.06.020

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 6000–6010.

Wang, W., & Li, B. (2024). Learning Personalized Privacy Preference from Public Data. *Information Systems Research*. https://doi.org/10.1287/isre.2023.0318

Westbrook, R. A., & Oliver, R. L. (1991). The Dimensionality of Consumption Emotion Patterns and Consumer Satisfaction. *Journal of Consumer Research*, *18*(1), 84–91. https://doi.org/10.1086/209243

Yu, L., & He, C. (2023). A shapelet-based behavioral pattern extraction method for credit risk classification with behavior sparsity. *Advanced Engineering Informatics*, *58*, 102227. https://doi.org/10.1016/j.aei.2023.102227

Žabkar, V., Brenčič, M. M., & Dmitrović, T. (2010). Modelling perceived quality, visitor satisfaction and behavioural intentions at the destination level. *Tourism Management*, *31*(4), 537–546. https://doi.org/10.1016/j.tourman.2009.06.005

Zhang, H., Liu, R., & Egger, R. (2024a). Unlocking Uniqueness: Analyzing Online Reviews of *Airbnb Experiences* Using BERT-based Models. *Journal of Travel Research*, *63*(7), 1688–1708. https://doi.org/10.1177/00472875231197381

Zhang, X., Zhang, Y., Long, D., Xie, W., Dai, Z., Tang, J., Lin, H., Yang, B., Xie, P., Huang, F., Zhang, M., Li, W., & Zhang, M. (2024b). *mGTE: Generalized Long-Context Text Representation and Reranking Models for Multilingual Text Retrieval* (arXiv:2407.19669). arXiv. http://arxiv.org/abs/2407.19669