SILESIAN UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE

SCIENTIFIC PAPERS OF SILESIAN UNIVERSITY OF TECHNOLOGY 2024
ORGANIZATION AND MANAGEMENT SERIES NO. 209

ASSESSING THE QUALITY OF AI-GENERATED CONTENT IN TOURISM - A SIGNALLING THEORY APPROACH

Marcin OLSZEWSKI

Department of International Economics, Poznań University of Economics and Business;
marcin.olszewski@ue.poznan.pl, ORCID: 0000-0001-9041-9953

Purpose: The study highlights gaps in AI research from the perspective of non-business users. The main aim of this article is to identify quality dimensions and signals used to evaluate the results generated by GenAI in tourism planning.
Design/methodology/approach: The study employed a qualitative research methodology, specifically an exploratory focus group. Conducted in Poland in 2024, the study assessed how Generation Z representatives search for quality signals in artificial intelligence-generated content (AIGC).
Findings: Using signalling theory, the signals considered in the evaluation process were identified and categorized. The study sheds light on the impact of visible signals such as links, images, and AI brands on AIGC satisfaction and how these signals relate to content quality dimensions. The findings indicate that relevant and accurate links are crucial as they help verify the information's authenticity, while irrelevant or incorrect links erode trust. Photographs enhance the plausibility of the answers and aid in visualizing attractions, though consistency and truthfulness remain vital to the quality of the response. The AI's brand, however, is deemed unimportant.
Originality/value: This paper adds to the developing field of tourism planning by focusing on the use of AIGC. Moreover, it contributes to the field of human-AI interactions, and human-centred AI, by investigating how users evaluate the quality of AIGC. The framework employs a unique approach by combining signalling and screening theory to analyze AIGC and elucidate the relationship between signals and perceived quality.
Keywords: artificial intelligence; signalling theory; AIGC; tourism; human-AI interactions.
Category of the paper: Research paper.

1. Introduction

As generative artificial intelligence (GenAI) is a recent technological development, there is a limited amount of research available on how customers respond to it. The focus of AI research remains on technological advancement, with the efficiency of these systems primarily evaluated through system performance metrics rather than the quality of their human interaction

capabilities (Jin, Zhang, 2023; Raees et al., 2024). The importance for the tourism industry of understanding customer responses to GenAI has been highlighted by Kim et al. (2023). They have also identified research gaps, including the absence of studies on customers' reactions to AI-generated content (AIGC).

Implementing AI in the travel planning and organization process has been found to have several benefits, including the objectivity of the results (Christensen et al. 2024), reduction of online information overload, and enhanced personalization (Kim et al., 2023). However, it is important to acknowledge that AI results may be subject to bias. The issue of AI hallucinations, i.e. making up results, is a concern that needs to be addressed to maintain trust in AI work. According to Wei et al. (2022), the reliability (truthfulness) of responses is the main factor determining their quality. Additionally, when planning and organizing holidays with AI, factors such as completeness, timeliness, and usefulness of the answer may also be considered (Kim et al., 2023). It is worth noting that the criteria used to evaluate performance may not always be clear, especially for younger individuals. According to Christensen et al. (2024), a significant number of Gen Z and Millennial consumers have shown a preference for the ChatGPT travel itinerary, even though it may contain inaccurate or fabricated information. This is due to their perception of ChatGPT as a more reliable content source than conventional and popular sources used for travel and tourism decision-making. In line with signalling theory, the paper assumes that, in the presence of information asymmetry, AIGC recipients search for observable cues indicating high or low quality content. Thus, it is important to identify not only the dimensions of AIGC quality, but also the signals that are perceived as proxies for these dimensions.

Taking this into consideration, the main aim of this article is to identify quality dimensions and signals used to evaluate the results generated by GenAI in tourism planning. This paper draws on two distinct bodies of literature. Firstly, it contributes to the emerging field of tourism planning, with a particular focus on the use of AI-generated content (Chen et al., 2023). Secondly, this study makes a contribution to the field of human-AI literature more broadly, and to the specific area of human-centred AI (Raees et al., 2024), by examining how users assess the quality of AI-generated content.

The potential for improving the quality of AI outputs has significant implications for AI content creators. Furthermore, the study provides valuable insights into consumer reactions and behaviours associated with AI interactions. The framework's unique approach combines signalling and screening theory to understand AI-generated content and clarify the relationship between signals and perceived quality.

2. Quality of AI-generated content

The investigation of content quality and its influence on consumer decision-making in the context of tourism has been a long-standing area of research (Z. Zhang et al., 2016). The focus of research has been on both marketer-generated information and consumer-generated information (Bickart, Schindler, 2001; Chua, Banerjee, 2016).

Artificial intelligence-generated content (AIGC) results from the process of creating digital content, including images, music, and natural language, through the use of AI models (Wang et al., 2023). The quality of AI-generated content (AIGC) and the consumer experience with AI are new areas of research that are becoming increasingly popular in service research (Chen et al., 2023; Khan, Mishra, 2023). Recent studies suggest that AI-generated content shares many attributes with content created by humans, but it tends to be more detailed, accurate, and effective (Zhang et al., 2024). The impact of consumers' interactions with AI tools can be either positive or negative. On the one hand, the advantages of use, such as convenience, speed of operation, and so forth, can lead to a positive experience. Conversely, negative experiences may result from factors such as hallucinations, vague data, and data security concerns.

The concept of content quality is related to the general idea of the quality of services understood as the ability to meet or exceed customer expectations (Parasuraman et al., 1986). Definitions of AIGC quality, therefore, focus on meeting the AI users' needs and requirements. AIGC typically comprises two phases: the first is the extraction and understanding of user intent information, while the second is the production of desired content based on the extracted intentions (Wang et al., 2023). The quality of the outcome is contingent upon the extent to which the generated results align with the user requirements initially specified in the prompt.

3. Dimensions of AIGC quality

Various researchers have studied the factors that influence users' intentions to use AI-based technology in tourism. According to Pillai and Sivathanu (2020), perceived ease of use, usefulness, and trustworthiness are among the significant factors. Melián-González et al. (2019) also found that expected performance is crucial. Loureiro et al. (2021) discovered that a tourist's perceived value impacts the quality of the relationship between tourists and intelligent voice assistants. Pham et al. (2024) conducted a study using three human-like cues - perceived warmth, speed of communication, and perceived competence. These cues stimulated cognitive responses, such as trust in ChatGPT and attitudes towards ChatGPT, resulting in increased satisfaction and intention to continue using ChatGPT for travel services.

The research on content quality indicates various ways of defining the dimensions of quality. In one of the pioneering studies on service quality, Grönroos (1984) formulated an early multi-dimensional framework for assessing service quality, identifying two main dimensions: technical quality and functional quality. The technical quality of AI-generated content is contingent upon the substance of the response, which can be described by the relevance of the response, detail, veracity, and other content-related dimensions. In contrast, the functional quality of AI-generated content is concerned with the form and delivery of the response. With regard to the functional quality of AI-generated content, the aesthetic presentation, graphical form, and speed of generation of results are of particular importance.

On the other hand, the quality of information can be understood through three dimensions: comprehensibility, specificity, and reliability (Chua, Banerjee, 2016; Kim, Lei, 2024). The comprehensibility of AI-generated content can be measured by determining how easily it can be understood (Fang et al., 2016). The specificity of text created by AI can be assessed by evaluating the amount of relevant information it provides to assist with decision-making. The reliability of information can be gauged by determining the extent to which consumers trust it (Table 1).

Table 1.
AIGC quality dimensions

Quality dimension	Description	Source
Reliability	Absence of false information	Chua, Banerjee (2016); Kim, Lei (2024)
Comprehensiveness	Specificity and information depth, relevance to the purpose	Chua, Banerjee (2016); Kim, Lei (2024)
Readability	Ease of understanding, simplicity of language	Chua, Banerjee (2016); Fang et al., (2016)

Source: Authors own work.

The question of what users perceive as quality, regardless of how it is defined, is an important one. This study employs signalling and screening theory (Spence, 2002) to address the issue of proxies of the quality of AI-generated content in tourism planning. Signaling theory, as introduced by Michael Spence, originally emerged in the context of labor markets to explain how individuals convey their abilities or qualifications to potential employers. In this framework, "signals" are observable attributes or actions that convey information about some unobservable quality.

The signalling theory suggests that in the presence of an information imbalance, the party with less information will seek to employ different information signals or indicators to bridge the gap and enhance their decision-making process (Connelly et al., 2010). According to Pemer and Skjølsvik (2019), this theory explores the process by which receivers interpret and evaluate signals from signalers to gain insight into their quality.

It is presumed that the quality of AI-generated content is unknown to the users. It is only when users assess the output generated by the AI that they can conclude the quality of the result. The act of seeking out quality information can be considered a form of screening.

Similarly, the sending of information (or cues) that suggest high quality can be considered a form of signalling. In light of the above, the article seeks to address the following questions: what signals do GenAI users use to identify and differentiate between high and low quality content in the tourism planning process?

4. Method

In 2024, a study was conducted in Poland to examine the process of searching for quality signals in response, based on the views of representatives from Generation Z.

The study employed a qualitative research method, specifically an exploratory focus group. In this study focus group was „a carefully planned discussion designed to obtain perceptions on a defined environment" (Kreuger, 1998 p. 88 in Smithson, 2000).

This approach is conducive to the generation of novel ideas that emerge within a social context. One of the key benefits of this methodology is that it allows research participants to collaborate and develop ideas collectively, ensuring that their priorities and perspectives are taken into account (Smithson, 2000). This approach helps to create a theory firmly rooted in the actual experiences and language of the participants. The use of focus groups is not without limitations. One such limitation is the tendency for socially acceptable opinions to emerge, as well as the tendency for certain types of participants to dominate the research process (Smithson, 2000).

Individuals meeting the following inclusion criteria were invited to participate in the study: representatives of Generation Z who use generative artificial intelligence and are willing to participate. According to Breen, (2006), theoretical saturation in focus-group research is normally reached after 10–12 interviews. In this study, the group included 18 individuals. The subjects involved were between 20 and 22 years of age and were mostly male (n = 10).

The research procedure was based on Smithson's (2000) and Breen, (2006) recommendations and involved three stages:

- In the first stage, participants were requested to search for content using leading GenAIs such as ChatGPT by OpenAI, Gemini by Google, and Copilot by Microsoft for trip inspiration, destination information, and details on dates, prices, and service providers.
- In the second stage, participants were requested to provide an evaluation of the results they obtained. They were asked to indicate whether they were satisfied or dissatisfied with the results and to explain the reasons behind their choice. Additionally, respondents were asked to specify the criteria by which they evaluated the quality of an answer as either positive or negative.

- In the third stage, the results were transcribed and analysed using Atlas.ti. The respondents' answers were categorised by two researchers independently into dominant themes, which included both AIGC quality dimensions and observable signals.

The objective of the study was to determine how users form opinions about the quality of AI-generated content. The user responses were subjected to an analysis to identify any observable signals that might indicate a positive or negative sentiment.

5. Results

The analysis of the results commenced with the presentation of a word map (Figure 1), constructed from the responses provided during the survey.

Figure 1. Word cloud based on user responses on AICG quality results.

Source: Authors own work.

The participants in the study were asked to indicate which of the GenAI results met their expectations the most. The itinerary presented by OpenAI's ChatGPT received the highest score in the evaluation by participants. Additionally, 20% of participants indicated that the Microsoft Copilot performed best in preparing the itineraries, while none of the respondents felt that the Google Gemini itineraries were the most optimal.

Subsequently, the key cues that respondents considered when evaluating the AIGC were identified through analysis of the responses. This part of the survey involved two people mutually controlling how the responses were classified. The responses were classified into four main categories, which serve as key indicators (signals) of AIGC quality:

- Length and Depth of Response.
- Use of Links.
- Use of Photos.
- Brand of the GenAI.

In the next step, signals were linked to quality attributes (Chua, Banerjee, 2016; Kim, Lei, 2024). The results are presented in Figure 2.

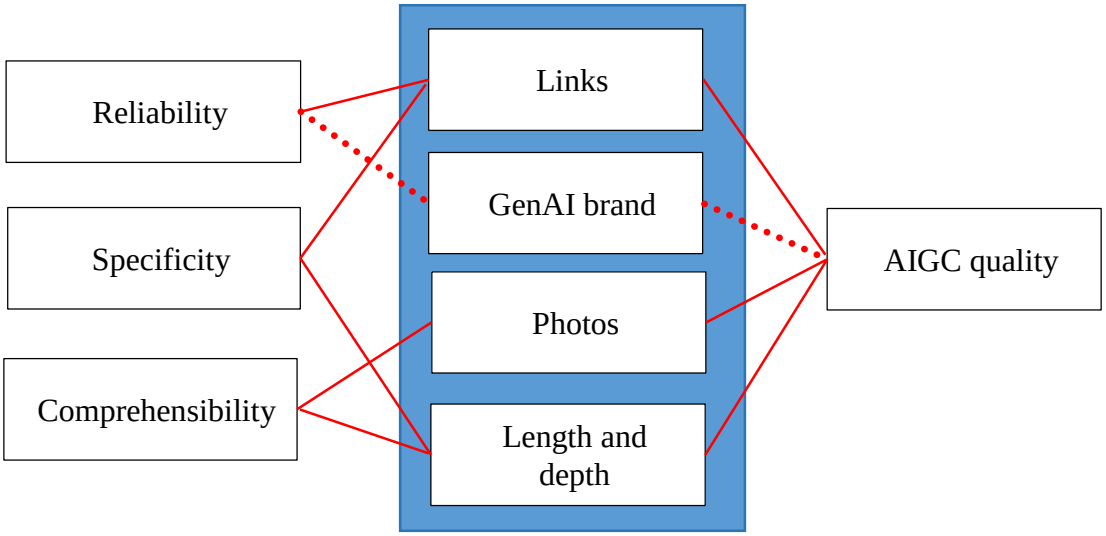

Figure 2. Word cloud based on user responses.

Source: Authors own work.

The length and depth of the results are considered proxies for the quality of the results. On the one hand, longer and more detailed answers allow for more knowledge and correspond better to the need to know a place. On the other hand, shorter answers are easier to read and absorb.

As the participants in the interviews indicated:

- „The length and depth of the answer can indicate quality".
- „The clearest, simplest answer that looks credible".
- „A good response avoids long, convoluted sentences and accumulates information effectively".

Another signal that users employ to infer quality is the presence of images that complement the text of the responses. As indicated by the participants:

- „Photos can influence interest in a place by showing that it exists and making the information more engaging".
- „Photographs made the answer seem credible, they can influence interest in a particular object/place".
- „However, photos are not always necessary, and their importance can vary".

Additionally, relevant proxies of AIGC include links to external websites embedded within the text of the answer. The inclusion of such links, provided they are genuine and direct to actual external pages, serves to enhance both the credibility of the answers and their depth. As stated by the individuals who participated in the interviews:

- „The links were meaningful. They allowed me to verify whether AI's answers made any sense".
- „Links make a difference to the speed of verifying the information provided, they certainly make this easier, but if the app provides random links then it becomes immediately unreliable".
- „Links are helpful but not essential, and their presence can make it easier to check the provided information".

GenAI's branding may indicate a higher reliability of the responses. The majority of respondents indicated that the responses generated by ChatGPT, a tool developed by OpenAI, were of a notably high quality. OpenAI is a well-known and highly regarded brand in artificial intelligence. At the same time, participants emphasised that the content itself was more important than the GenAI brand.

As indicated by the participants:

- „In my opinion, the content is more important than the name".
- „The name does not matter more than the content".
- „The content of the answer is very important, but the brand can also indicate the sophistication of the artificial intelligence".

According to the survey, the role of generative AI is not only to provide content based on the user's query but also to ensure that it is presented in a way that gives the impression of high-quality content. In other words, it should appear reliable, readable and complete. To effectively evaluate content quality, it is important to consider multiple signals in conjunction. Relying on a single signal, such as the presence of links, can lead to biased or incomplete assessments.

6. Discussion

Fluctuations in AIGC quality present a significant challenge to the widespread application of AI in tourism activity planning. It is therefore essential to understand the performance evaluation process from the perspective of those using it and to develop mechanisms to identify low-quality content.

Signalling theory provides a valuable framework for understanding how content quality is perceived and evaluated. By examining the signals emitted by an AI platform - whether through links, images, content length and depth, or the AI brand - consumers can make more informed decisions about content quality. The challenge, however, is to distinguish authentic signals from

those that are manipulated or misleading, and to ensure that the mechanisms in place (both human and algorithmic) are sophisticated enough to adapt to the evolving nature of these signals.

The study presents findings on the factors that influence the use of AI in travel planning. It acknowledges research gaps in the area of AI research from a non-business user perspective. The signals considered in the evaluation process were identified and classified using signalling theory. This study provides insights into the impact of visible cues such as links, images and brand AI on AIGC satisfaction and how these signals relate to the dimensions of content quality. First, it was found that the inclusion of links is important as they help to verify the veracity of information. Links should be relevant and accurate; random or incorrect links undermine trust. Photographs are also helpful as they make the answer more plausible and help to visualise the attractions. However, consistency and truthfulness are critical to the quality of the answer. The name of the AI is not considered important.

The study offers practical recommendations for the development of GenAI tools dedicated to tourism planning. To create quality content, it must first meet three criteria: reliability, specificity and comprehensibility. However, this is not the full picture. In addition, it is necessary to understand how to communicate these qualities by using signals and cues from which users will infer quality. Content should be free of hoaxes, of an appropriate length, supplemented with links and good-quality images. Furthermore, it should be free of errors and misleading information, and it should be presented in a clear and concise manner.

Further research is required to identify additional methods for assessing the quality of AIGC results. Moreover, it is recommended that research be conducted using quantitative response analysis methods, such as the Gunning-Fog Index, the Coleman-Liau Index, and the Automated Readability Index, among others (Chua, Banerjee, 2016). This will enable the objective assessment of AIGC quality attributes. Furthermore, research involving large samples of respondents is essential. The advancement of GenAI technology must be accompanied by the further development of research dedicated to the interactions between AI and humans, with a particular focus on human-centric research.

In light of the considerable capabilities of GenAI in the creation of content, it is imperative that humans engage in a process of critical examination and review of the results produced by this technology. It is thus vital to maintain a continuous examination of the process of evaluating and appraising AI-generated content. Future applications of signalling theory to the evaluation of AI-generated content in tourism could include the analysis of how well the content signals its quality and reliability to potential tourists. This could help to develop better AI systems for content generation and more effective ways for users to assess the trustworthiness of AI-generated travel information.

Acknowledgements

Supported by funding from the Ministry of Science (Poland) under the 'Regional Initiative of Excellence' programme.

References

1. Chen, C., Fu, J., Lyu, L. (2023). *A pathway towards responsible AI generated content*. arXiv (Cornell University), available at: https://doi.org/10.48550/arxiv.2303.01325
2. 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*, pp. 1-16.
3. Chua, A.Y., Banerjee, S. (2016). Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality. *Computers in Human Behavior, Vol. 54*, pp. 547-554.
4. Connelly, B.L., Certo, S.T., Ireland, R.D., Reutzel, C.R. (2010). Signaling Theory: A Review and assessment. *Journal of Management, Vol. 37, No. 1*, pp. 39-67.
5. Fang, B., Ye, Q., Kucukusta, D., Law, R. (2016). Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management, Vol. 52*, pp. 498-506.
6. González, S.M., Taño, D.G., Bulchand-Gidumal, J. (2019). Predicting the intentions to use chatbots for travel and tourism. *Current Issues in Tourism, Vol. 24, No. 2*, pp. 192-210.
7. Grönroos, C. (1984). A Service Quality Model and its Marketing Implications. *European Journal of Marketing, Vol. 18 No. 4*, pp. 36-44.
8. Jin, F., Zhang, X. (2023). Artificial intelligence or human: when and why consumers prefer AI recommendations. *Information Technology and People*.
9. Khan, A.W., Mishra, A. (2023). AI credibility and consumer-AI experiences: a conceptual framework. *Journal of Service Theory and Practice, Vol. 34, No. 1*, pp. 66-97.
10. Kim, B., Lei, S.I. (2024). Millennials' interactions with social media content on cruise tourism. *Tourism Review*.
11. Kim, J., Kim, J., Kim, C., Kim, S. (2023). Do you trust ChatGPTs? Effects of the ethical and quality issues of generative AI on travel decisions. *Journal of Travel & Tourism Marketing, Vol. 40, No. 9*, pp. 779-801.
12. Loureiro, S.M.C., Japutra, A., Molinillo, S., Bilro, R.G. (2021). Stand by me: analyzing the tourist–intelligent voice assistant relationship quality. *International Journal of Contemporary Hospitality Management, Vol. 33, No. 11*, pp. 3840-3859.

13. Parasuraman, A., Zeithaml, V.A., Berry, L.L. (1986). *Servqual: a multiple-item scale for measuring customer perceptions of service quality*. Marketing Science Institute eBooks, available at: https://ci.nii.ac.jp/ncid/BA22737010
14. Pemer, F., Skjølsvik, T. (2019). The cues that matter: Screening for quality signals in the ex ante phase of buying professional services. *Journal of Business Research, Vol. 98*, pp. 352-365.
15. Pham, H.C., Duong, C.D., Nguyen, G.T. (2024). What drives tourists' continuance intention to use ChatGPT for travel services? A stimulus-organism-response perspective. *Journal of Retailing and Consumer Services, Vol. 78*, p. 103758.
16. Pillai, R., Sivathanu, B. (2020). Adoption of AI-based chatbots for hospitality and tourism. *International Journal of Contemporary Hospitality Management, Vol. 32, No. 10*, pp. 3199-3226.
17. Raees, M., Meijerink, I., Lykourentzou, I., Khan, V., Papangelis, K. (2024). From explainable to interactive AI: A literature review on current trends in human-AI interaction. *International Journal of Human-computer Studies, Vol. 103301*.
18. Shin, D., Zhong, B., Biocca, F.A. (2020). Beyond user experience: What constitutes algorithmic experiences? *International Journal of Information Management, Vol. 52*, p. 102061.
19. Smithson, J. (2000). Using and analysing focus groups: Limitations and possibilities. *International Journal of Social Research Methodology, Vol. 3, No. 2*, pp. 103-119.
20. Spence, M. (2002). Signaling in retrospect and the informational structure of markets. *The American Economic Review, Vol. 92, No. 3*, pp. 434-459.
21. Wang, Y., Pan, Y., Yan, M., Su, Z., Luan, T.H. (2023). A survey on ChatGPT: AI-Generated Contents, Challenges, and Solutions. *IEEE Open Journal of the Computer Society, Vol. 4*, pp. 280-302.
22. Wei, H., Liu, W., Xie, J., Zhang, S. (2022). Social support to mitigate perceived risk: the moderating effect of trust. *Current Issues in Tourism, Vol. 26, No. 11*, pp. 1797-1812.
23. Zhang, J.J., Wang, Y.W., Ruan, Q., Yang, Y. (2024). Digital tourism interpretation content quality: A comparison between AI-generated content and professional-generated content. *Tourism Management Perspectives, Vol. 53*, p. 101279.
24. Zhang, Z., Zhang, Z., Yang, Y. (2016). The power of expert identity: How website-recognized expert reviews influence travelers' online rating behavior. *Tourism Management, Vol. 55*, pp. 15-24.