Artificial Intelligence and Sustainable Tourism Planning: A Hetero-Intelligence Methodology Proposal

Inteligencia Artificial y Planificación Turística Sostenible: Una propuesta de Metodología Hetero-inteligente

Eva María Buitrago-Esquinas
Universidad de Sevilla,Facultad de Ciencias Económicas y Empresariales, Departamento Economía Aplicada III, España, esquinas@us.es

Rocío Yñiguez-Ovando
Universidad de Sevilla,Facultad de Ciencias Económicas y Empresariales, Departamento Análisis Económico y Economía Política, España, ovando@us.es

Miguel Puig-Cabrera
Universidade do Algarve, Research Centre for Tourism, Sustainability and Well-being (CINTURS), Portugal, mpcabrera@ualg.pt

Margarida Custódio Santos
Universidade do Algarve, School of Management, Hospitality and Tourism (ESGHT); Research Centre for Tourism, Sustainability and Well-being (CINTURS), Portugal, mmsantos@ualg.pt

José António C. Santos
Universidade do Algarve, School of Management, Hospitality and Tourism (ESGHT); Research Centre for Tourism, Sustainability and Well-being (CINTURS), Portugal, jasantos@ualg.pt

Received: 15.11.2023; **Revisions required:** 22.02.2024; **Accepted**: 24.06.2024

Abstract

This study explores the growing significance of Large Language Models (LLMs) in tourism, for their current and potential applications. It aims to achieve two primary objectives: first, to develop a novel hetero-intelligence framework merging human and artificial intelligence (AI) to address contemporary sustainability challenges in tourism; second, to validate this framework by applying it to sustainable tourism planning, assessing LLMs' capabilities and limitations. The research employs a hetero-intelligence performance test, contrasting human intelligence and AI contributions in sustainable tourism planning with overtourism as a proxy challenge. Results showed that hetero-intelligence could effectively address sustainability issues in tourism, provided human and AI strengths and weaknesses are understood. LLMs proved useful in diagnosing and proposing solutions for sustainability-related issues. However, a rigorous methodological framework is essential to ensure unbiased outcomes. The research offers practical guidelines for applying this approach and significantly contributes to epistemological and empirical dimensions, providing valuable insights for researchers and tourism planners. The study calls for more empirical research to validate the methodology and explore ethical and legal dimensions, extending hetero-intelligence applications to broader sustainability challenges in tourism.

Keywords: Sustainable tourism planning; Hetero-intelligent performance testing; Human intelligence; Conversational generative AI; Large language models; ChatGPT .

Resumen

Este estudio explora la creciente importancia de los Modelos de Lenguaje Grande (LLMs) en el turismo en sus aplicaciones actuales y potenciales. Se plantean dos objetivos principales. Primero, desarrollar un nuevo marco hetero-inteligente que, combinando la inteligencia humana y artificial (IA), aborde los desafíos actuales de sostenibilidad en el turismo; segundo objetivo, validar este marco para la planificación turística sostenible, evaluando capacidades y limitaciones de los LLM. La investigación emplea una prueba de desempeño hetero-inteligente, contrastando las contribuciones de la inteligencia humana y la IA en la planificación del overtourism. Los resultados mostraron que la hetero-inteligencia podría abordar eficazmente los problemas de sostenibilidad en el turismo, teniendo en cuenta las fortalezas y debilidades de los humanos y de la IA. Los LLM resultaron útiles para diagnosticar y proponer soluciones a cuestiones sobre la sostenibilidad. Sin embargo, un marco metodológico riguroso es esencial para garantizar resultados imparciales. La investigación proporciona conocimientos valiosos para investigadores y planificadores turísticos. Pero es necesario mayor investigación empírica para validar la metodología y explorar dimensiones éticas y legales en desafíos más amplios de sostenibilidad en el turismo..

Palabras clave: Planificación turística sostenible; Hetero-inteligencia; Inteligencia humana; Inteligencia artificial generativa conversacional; Modelos de lenguaje grandes; ChatGPT.

1. Introduction

A broad consensus has been reached on the disruptive nature of large language models (LLMs) as a new approach that could change societies, personal interactions and the skills, abilities and knowledge individuals need for optimal job performance in the near

future (Dwivedi et al., 2023). LLMs are 'artificial intelligence (AI) tools based on multi-layer recurrent neural networks that are trained on vast amounts of data to generate human-like text' (Alberts et al., 2023, p. 1549). These models diversify and evolve within each task performance, so the interactions between humans and AI create new scenarios that demand cooperation between both intelligences to produce change in society, and which we propose to call a hetero-intelligence system. Such a system incorporates, on the one hand, the concept of hybrid intelligence, which is defined by Dellermann et al. (2019, p. 3) as 'systems that have the ability to accomplish complex goals by combining human and artificial intelligence to collectively achieve superior results than each of them could have done in separation and continuously improve by learning from each other'. On the other hand, in order to align it with a social sciences framework, we apply Breakspear's dimension of intelligence as a capability to forecast change in time involving foresight and insight in a given context to provide an adequate solution to the identified challenge (Breakspear, 2013). Thus, adopting Breakspear's dimension of intelligence, we propose the concept of hetero-intelligence, which could be understood as a mixed capability of human and artificial intelligences that, combined, forecasts change in time involving foresight and insight in a given context to provide an adequate solution to the identified challenge. Furthermore, we propose that the process be conducted under human supervision in order to reach its full potential.

LLMs are attracting attention in multiple sectors, such as these models' potential for making medical diagnoses (Biswas, 2023) or even writing a medical essay (Eggmann et al., 2023). Other fields are also evaluating this technology's utilities, so research on AI is expanding in many areas. The literature on LLMs in tourism is still in an early stage given that most papers are theoretical or opinion-based, focusing until now on implications for the tourism industry, tourists' consumer behaviour (Carvalho & Ivanov, 2024; Dwivedi et al., 2023, 2024; Gursoy et al., 2023; Mich & Garigliano, 2023) and tourism education and research (Iskender, 2023; Ivanov & Soliman, 2023; Skavronskaya et al., 2023; Ülkü, 2023). These publications contain few references to tourism and sustainability, limiting themselves to pointing out examples of potential LLM applications, such as tourism flow management (Carvalho & Ivanov, 2024; Dwivedi et al., 2023). Scholars have thus reached a broad agreement about the need for future research agendas that delve deeper into this technology's challenges and opportunities for both more sustainable tourism (Majid et al., 2023; Mich & Garigliano, 2023; Rodríguez et al., 2023) and better planning (Dwivedi et al., 2023, 2024). The literature on LLMs' potential within tourism is growing quickly (Carvalho & Ivanov, 2024; Dwivedi et al., 2024; Henriques et al., 2024; Lacárcel, 2022; Zhang & Prebensen, 2024), but there is a lack of methodological and empirical studies to advance the understanding of the topic (Dwivedi et al., 2024; Gursoy et al., 2023; Izzo & Picone, 2022). Additionally, there is a significant gap concerning the potential impact of LLMs on tourism planning and sustainability. A review of the relevant literature revealed no empirical research on LLMs' capabilities and limitations as a sustainable tourism planning tool.

To address the identified gaps, the purpose of the present study is twofold: to propose an exploratory hetero-intelligence methodological framework based on the co-performance of human intelligence and AI that provides ad hoc expertise for diagnoses and policy-making focused on the main sustainable challenges in tourism. Also, the proposed framework will be tested using a specific sustainable tourism planning challenge in order to assess LLMs' capabilities and limitations in this context. Overtourism was chosen as a proxy to explore these models' potential as a sustainable tourism planning tool and to provide empirical evidence of their output's environmental, social and economic impacts on destinations (Yrigoy et al., 2023). More specifically, the current research concentrated on the case of Venice as one of the most widely recognised instances of overtourism in the literature (Celata & Romano, 2022; Goodwin, 2021; Torres, 2021). The rest of this paper is organised as follows. Section two provides a literature review. Section three is devoted to the methodology applied, while section four presents the results, including an in-depth discussion. Finally, section five presents the conclusions and implications.

2. LLMS' capabilities and limitations: review of the literature on tourism

These models' disruptive nature has stimulated a new debate about their implications for society, economies and academic fields. These impacts could vary greatly depending on the area involved. An extensive review of the literature was conducted to develop a systematic overview of Chat Generative Pre-trained Transformer (ChatGPT) and similar LLMs technologies' main capabilities and shortcomings concerning tourism, in general, and tourism planning, in particular. This review took as its starting point Dwivedi et al.'s (2023) study, which offers a multidisciplinary approach to LLMs based on 43 experts' work in different subject areas. The findings of the cited research were deepened by focusing on the literature, specifically on tourism and generative and conversational AI.

2.1 *LLMs' capabilities*

As shown in Table 1, the first group of capabilities is related to the information that these models can retrieve and process. Technologies such as ChatGPT can gather and summarise a large amount of data in a short time at an acceptable level of quality (OpenAI, 2023). LLMs are capable of generating narrative texts and a wide range of new content (i.e. images, videos and software) that can be used in reports, emails, blog posts and social networks, among many other applications. These models can also recover quantitative and qualitative data from the databases in which these tools are trained.

Table 1- LLMs capabilities: literature review

LLMs Capabilities	References
C1. Process and summary of large amounts of information in a short time and acceptable quality: C1a. Generation of texts and feeds C1b. Assessment of different topics C1c. Simplification of information searching	Carvalho & Ivanov (2024); Dwivedi et al. (2023, 2024); Gursoy et al., 2023; Ivanov & Soliman (2023); Mich & Garigliano (2023); Paul et al. (2023); OpenAI (2023); Sifat (2023)
C2. Recovery of existing quantitative and qualitative information	Carvalho & Ivanov (2024); Dwivedi et al. (2023, 2024)
C3. Capacity (limited) for analysis, interpretation, and prediction: C3a. Preliminary analysis C3b. Suggestions and preliminary statements C3c. Trends identification	Carvalho & Ivanov (2024); Dwivedi et al. (2023, 2024), Mich & Garigliano (2023); Sifat (2023); Paul et al. (2023)
C4. Interaction with users offering sophisticated answers: C4a. Doubt solution and specific support	Carvalho & Ivanov (2024); Dwivedi et al. (2023, 2024); Gursoy et al. (2023); Ivanov & Soliman (2023); Mich & Garigliano (2023); Paul et al. (2023)
C5. Brainstorming with users: C5a. Critical thinking fostering C5b. Creativity fostering	Carvalho & Ivanov (2024); Dwivedi et al. (2023); Ivanov & Soliman (2023); Mich & Garigliano (2023); Sifat (2023)
C6. Communication of ideas and self-awareness on several topics	Carvalho & Ivanov (2024); Dwivedi et al. (2023, 2024); Ivanov & Soliman (2023)
C7. Self-recognition of limitations	Dwivedi et al. (2023); OpenAI (2023); Paul et al. (2023)
C8. Accessibility, affordability and availability	Carvalho & Ivanov (2024); Gursoy et al. (2023); Mich & Garigliano (2023); OpenAI (2023); Sifat (2023)

Source: Own elaboration.

Researchers have documented how these information-related abilities can be useful in the tourism sector to personalise tourists' experiences before, during and after trips and help tourism companies complete a wide range of tasks in various areas. For example, LLMs are able to extract keywords or positive and negative aspects of reviews that can be used by tourism marketing departments (Carvalho & Ivanov, 2024). In addition, based on each tourist's preferences, these models are capable of identifying unique events, restaurants and hotels that meet special needs, thereby personalising experiences and increasing tourists' satisfaction (Dwivedi et al., 2023; Gursoy et al., 2023; Mich & Garigliano, 2023; Wong et al., 2023). However, no studies have focused on tourism planning, with only a few investigations giving examples of potential applications of information generated by LLMs for regulators seeking to monitor and control compliance with regulations and standards (Dwivedi et al., 2024).

The literature also reveals that ChatGPT and similar LLM technologies have capabilities that are still limited in terms of analysis, interpretation, and prediction. These tools merely report preliminary findings or statements based on the types of data with which they have been trained, which could still have useful applications in tourism (Carvalho & Ivanov, 2024; Dwivedi et al., 2023, 2024; Mich & Garigliano, 2023; Sifat, 2023). Mich and Garigliano (2023) specifically discuss how ChatGPT can be used in e-tourism to identify target markets.

LLMs can further interact with users to give them sophisticated answers to questions similar to those other human beings would provide. This capacity is one of the most studied uses within tourism, especially LLMs' integration into customer service (Carvalho & Ivanov, 2024; Dwivedi et al., 2023, 2024; Gursoy et al., 2023; Mich & Garigliano, 2023; Paul et al., 2023). These models can thus offer hyper-personalised recommendations to tourists and contribute to their empowerment (Dwivedi et al., 2024). Nonetheless, research on these capabilities' application to planning has been insufficiently disseminated.

Another LLM capability related to interactions with humans is these tools' brainstorming functions that can be used with any topic and task to foster critical thinking and creativity and develop new ideas for dealing with different challenges (Dwivedi et al., 2023; Mich & Garigliano, 2023). Scholars have discussed how these functions could be useful to tourism research (Ivanov & Soliman, 2023), including the creation of new products or experiences (Carvalho & Ivanov, 2024) or even policy-making (Sifat, 2023), but the literature still shows a void in terms of how LLM brainstorming can be integrated into planning to achieve more sustainable tourism. In addition, LLM-generated feeds and conversational flows are able to communicate ideas and raise awareness about varied issues for users, such as problems related to tourism (Carvalho & Ivanov, 2024; Dwivedi et al., 2023, 2024; Ivanov & Soliman, 2023). No analyses have focused on LLMs' possible positive effects on tourism sustainability, although potential uses have been found, including, among others, informing tourists – and raising their awareness – about sustainable practices (Dwivedi et al., 2024). This application could be an especially important way to expand travellers' knowledge about the need for sustainable tourism.

Finally, these tools are highly accessible, affordable, and available, given that LLMs have 24/7 capabilities in an increasing number of languages and technological devices (OpenAI, 2023). Studies have explored how LLMs can affect tourists' behaviour and the tourism industry but have failed to consider tourism planning. Gursoy et al. (2023) focused on these models' ability to optimise travel experiences by eliminating barriers such as language and facilitating more inclusive tourism. Carvalho and Ivanov (2024) also discuss how LLMs' affordability could contribute to the democratisation of AI because they are easily incorporated into small and medium-sized enterprises' tourism operations.

The latter idea can be extended to tourism planning at the lowest administrative levels (i.e. small municipalities), lacking extensive financial and human resources that could be partly made up for by using generative AI. All the LLM capabilities analysed have especially significant potential impacts on so-called knowledge work (Dwivedi et al., 2023), so the repercussions for tourism planners are clearly positive. The present study concentrated on meeting the challenge of defining these repercussions in order to improve sustainable tourism planning.

2.2 LLMs' limitations

Despite generative and conversational AI's high potential, the literature also identifies limitations that affect these technologies' performance depending on the tasks involved and efforts made to manage them properly (see Table 2). Multiple areas of knowledge have been evaluated in terms of ChatGPT usage. A common limitation found is shallow information in generated texts and content (Dwivedi et al., 2023), revealing that LLMs have trouble understanding complex technical concepts. More specifically, this finding implies seriously imprecise or even harmful instructions (Carvalho & Ivanov, 2024; Ivanov & Soliman, 2023; Mich & Garigliano, 2023). Researchers have also detected a lack of depth and difficulties in generating information related to specific contexts (Carvalho & Ivanov, 2024; Sifat, 2023). In tourism, Mich and Garigliano (2023) point out that, although ChatGPT reports an acceptable quality of information, many tasks cannot be accomplished with 'almost correct data', such as producing timetables or prices, so this information has to be verified by humans.

Table 2- LLMs Limitations: literature review

LLMs Limitations	References
L1. Lack of depth level of outputs on the conversational flow: L1a. Difficulties in understanding complex concepts L1b. Lack of precision L1c. Omissions L1d. Biases	Carvalho & Ivanov (2024); Dwivedi et al. (2023, 2024); Ivanov & Soliman (2023); Mich & Garigliano (2023); OpenAI (2023)
L2. Lack of specification and adaptation on outputs of the conversational flow: L2a. Concerns about understanding specific contexts L2b. Generalisations L2c. Biases	Carvalho & Ivanov (2024); Dwivedi et al. (2023); Sifat (2023)
L3. Limited social and cognitive skills: L3a. Limited interpretation of emotions, sentiments, and preferences L3b. Critical thinking L3c. Creative thinking L3d. Adaptation to specific contexts or situations	Carvalho & Ivanov (2024); Dwivedi et al. (2023); Gursoy et al. (2023); Iskender (2023); Sifat (2023); OpenAI (2023)
L4. Limited capacity to execute actions or make decisions: L4a. New data generation L4b. Negotiation with stakeholders L4c. Implementation of public policies L4d. Others	Dwivedi et al. (2023); Ivanov & Soliman (2023); OpenAI (2023)
L5. Limited interpretative, analytical, and predictive skills in current or future scenarios	Dwivedi et al. (2023); Iskender (2023);
L6. Technical limitations: L6a. Lack of update L6b. Lack of transparency L6c. Social bias L6d. Not reference sources L6e. Cybersecurity	Carvalho & Ivanov (2024); Dwivedi et al. (2023); Gursoy et al. (2023); Ivanov & Soliman (2023); OpenAI (2023); Paul et al. (2023); Sifat (2023)
L7. Ethical, legal, and socioeconomic implications	Carvalho & Ivanov (2024); Dwivedi et al. (2023, 2024); Gursoy et al. (2023); Iskender, 2023, Ivanov & Soliman (2023); Sifat (2023)

Source: Own elaboration.

Another significant group of limitations is related to social and cognitive skills (Dwivedi et al., 2023). Technologies such as ChatGPT have particular difficulty interpreting users' emotions, feelings, sentiments and preferences (Gursoy et al., 2023; Sifat, 2023), especially in terms of these tools' inability to adapt to changing environments and restricted critical and creative thinking (Iskender, 2023). In addition, LLMs by nature cannot execute actions or make decisions in the real world. Scholars who have explored specific areas of knowledge have remarked on LLMs' limited ability to generate new data (e.g. formulate indicators or surveys) (Ivanov &

Soliman, 2023), negotiate with stakeholders or implement policies (Dwivedi et al., 2023). These models can analyse and predict, but their functions are still too limited to be considered critical analysis (Iskender, 2023) and reliable predictions of the future (Dwivedi et al., 2023).

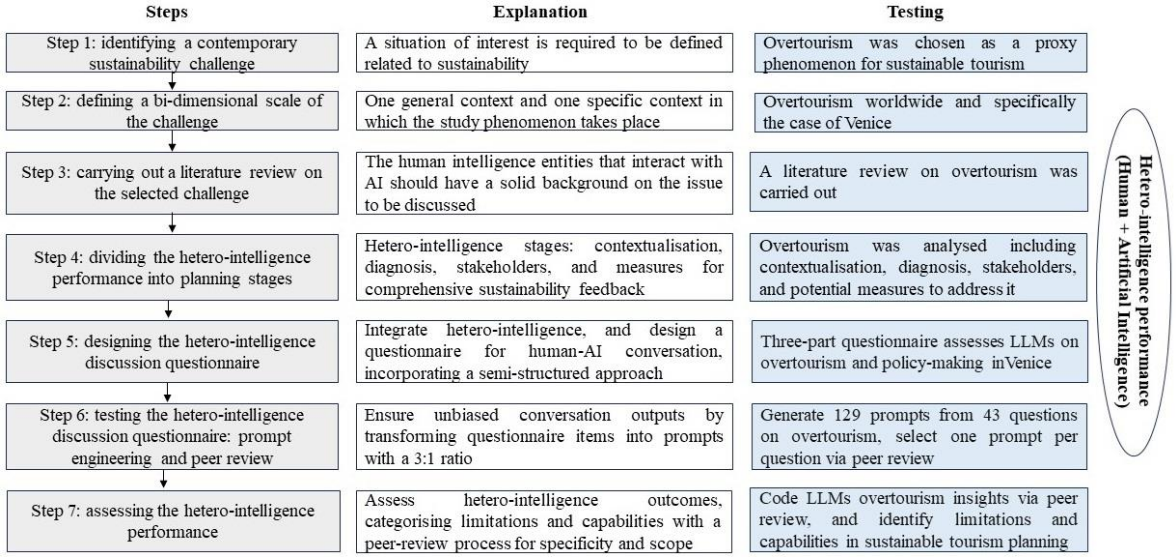
Finally, a significant series of technical limitations could endanger these technologies' reliability, including that of ChatGPT (Dwivedi et al., 2023; Paul et al., 2023), by generating serious concerns about cybersecurity and legal and ethical issues still under debate. First, ChatGPT's output contains bias related to the information used to train this model. This shortcoming is aggravated by a lack of transparency about the data and algorithms used for training and the omission of specific sources used to generate output for each prompt. To address these biases, greater transparency must become an essential part of the optimal evolution of LLMs. Second, the absence of updates is also a limitation since ChatGPT 4.0 only contains information until September 2021 (OpenAI, 2023). These technical limitations and their implications have been discussed in the tourism literature (Carvalho & Ivanov, 2024; Dwivedi et al., 2023, 2024; Gursoy et al., 2023; Iskender, 2023), which focuses on the risk of increasing the digital divide, on the impacts on labour markets or on the possible decrease in trips' social value due to tourists' potential isolation. This literature has not yet specifically analysed how LLMs' limitations may affect tourism planning, but all authors agree these shortcomings should be highlighted for management as one of the challenges that must be dealt with in tourism research and practice.

The above literature review confirmed that numerous gaps exist in studies on LLMs within tourism. More methodological and empirical research is needed, and scholars need to address these tools' specific problems regarding tourism planning and sustainability. Based on these models' capabilities and limitations previously analysed, the present study sought to fill these gaps by proposing and testing a hetero-intelligence methodology on the problem of overtourism in Venice, Italy.

3. Hetero-intelligence methodology for sustainable tourism planning

The integration of AI into the tourism sector constitutes a significant stride towards using innovative methods to make the most of technology's ability to contribute directly to policy-making processes. This research's pioneering methodology concentrated on contrasting artificial and human intelligence to guide conversational flow and assess how a hetero-intelligence approach can be integrated into sustainable tourism planning. The proposed hetero-intelligence method of sustainable tourism planning was built upon seven steps, each of which applied the main processes of other previously consolidated methodologies. This methodology's main pillars were based on a semi-structured interview involving a human and an AI-based entity (Buitrago-Esquinas et al., 2024). For this reason, all the steps had to be followed rigorously, as presented in Figure 1.

Figure 1- Methodological guidelines to apply a hetero-intelligent methodology for policy-making regarding a sustainable tourism challenge



Source: Authors.

Step one: Identifying a contemporary sustainability challenge. The situation to be analysed has to be defined and include a problem, issue or event (Crabolu et al., 2023) related to sustainable tourism and referred to by policy-makers. Overtourism was selected as a proxy for sustainable tourism planning challenges to test the proposed methodology as this issue is one of the biggest problems currently faced by the tourism sector.

Step two: Defining a bi-dimensional scale of the challenge. One of the main difficulties when testing the performance of AI-based entities as a tool supporting sustainable tourism planning is related to their capacity to understand specific contexts (Carvalho & Ivanov, 2024; Dwivedi et al., 2023; Sifat, 2023). To address this concern, the identified sustainable challenge must be framed using a bi-dimensional scale, at least one general and one specific context in which the phenomenon under study takes place. The present research examined overtourism on a global scale and a well-known case study of this phenomenon, namely Venice (Celata & Romano, 2022; Goodwin, 2021). In this way, the capabilities and limitations of AI and human intelligence interactions could be contrasted while focusing on a specific sustainable tourism challenge for a general and a specific context.

Step three: Carrying out a literature review on the selected challenge. The human intelligence that is going to interact with the AI has to have a solid background in the issue under discussion so that the hetero-intelligence performance test can be carried out successfully. In this case, the most updated research was collected to compose a proper state-of-the-art summary of overtourism in general and in the chosen case study (e.g. Venice). Fifty relevant papers were chosen to complement previously acquired knowledge about overtourism.

Step four: Dividing the hetero-intelligence performance into planning stages. This step involves defining the different stages in which the hetero-intelligence performance is going to be present within sustainable tourism planning. This process includes multiple stages, for example, contextualisation to understand the intrinsic factors and structure of the issue under study and diagnosing and monitoring to identify and/or analyse crucial data related to the problem within the previously defined bi-dimensional scale. Another possible stage is stakeholders' involvement in identifying different types of actors, defining their different interests, and participating in the chosen sustainable tourism challenge. A fourth possible stage is the definition of measures, decision-making, and implementation processes, including brainstorming measures that address sustainability issues and adapt these solutions to the specific contexts chosen. A 360º approach is achieved when all stages are considered to obtain more comprehensive feedback within the entire process, as was done to test the proposed methodology.

Step five: Designing the hetero-intelligence discussion questionnaire. This step involves developing a questionnaire based on a semi-structured interview (Adeoye-Olatunde et al., 2021; Çelik & Çevirgen, 2021). The questionnaire should include open-ended questions, as the absence of pre-defined response options will force the AI to provide detailed answers. The process includes pilot testing, content validation and triangulation (Flick, 2023). The questionnaire questions must be adapted to the bi-dimensional scale used, and all stages of sustainable tourism planning must be considered. The questionnaire should also be based on a literature review and the human experts' personal knowledge (Valentinas et al., 2022). For instance, this study's questionnaire was divided into three sections with a total of 43 open-ended questions. The first section was a self-diagnosis in which the selected LLM was asked about its capabilities and limitations within sustainable tourism planning. The second section focused on aspects related to diagnoses of overtourism. The last section was related to formulating policies to manage this problem.

Step six: Testing the hetero-intelligence discussion questionnaire: prompt engineering and peer review. Prompt design can condition the output obtained from human intelligence's conversation with an LLM. For this reason, each item of the current questionnaire should be transformed into a prompt to be tested by the chosen LLM. A 3:1 ratio is applied, thereby making two alternative questions in addition to the original item. In the present case, 129 prompts were made out of the 43 original questions. The main goal of this step is to choose 1 prompt out of 3 for each question and discard the alternative versions of that question. This prompt engineering process (Busch et al., 2023; Lo, 2023) had been previously validated by a peer review conducted by the researchers involved in the current research, which consisted of each reviewer having the same conversation in parallel with an LLM to select 1 out of 3 prompts per question. Finally, the authors discussed the results so that the 43 final prompts were chosen as those producing the most complete output from the selected LLM. The sixth step ends with the finalised version of the questionnaire that will guide the conversation between human intelligence and AI (See Annex 1).

Step seven: Assessing the hetero-intelligence performance. The last step is to categorise the outcomes of hetero-intelligence conversations into limitations and capabilities, which are assessed with reference to the specifics and scope inherent in the prompts used. This procedure requires the research team to conduct another peer-review process that classifies each prompt's output as a general versus specific capability or limitation. The identification of limitations and capabilities is a common way to work with LLMs in varied tasks (Dwivedi et al., 2023) so that the human experts' academic-practical knowledge can complement and contrast with the AI results regarding the issue under study. In the current case, overtourism phenomena were discussed with a ChatGPT-4 version developed by OpenAI because of this tool's similarities with other LLMs (Alipour et al., 2024). First, the information obtained with each question was coded according to the general limitations and capabilities of LLMs systematised by the researchers in tables 1 and 2 (discussed in section 2. The coding was carried out based on a peer-review process done independently by five researchers (see Annex 2). Second, this coding identified ChatGPT -4's specific capabilities and shortcomings for each task involved in sustainable destination planning. Last, the different intelligences' (i.e. artificial and human) roles in tourism planning were contrasted, thereby clarifying the skills humans must develop to complement LLMs' limitations and capabilities to maximise what AI can offer in this area.

4. Results and discussion: CHATGPT's capabilities and limitations within sustainable tourism planning

This section provides a deeper analysis and discussion of the present test of hetero-intelligence performance in sustainable tourism planning, with overtourism as a proxy planning challenge. The findings are presented and examined to contrast human intelligence and AI's contributions to the completion of specific tasks in planning processes, namely, contextualisation, diagnosis, monitoring, stakeholder involvement analysis, and policy- and decision-making procedures. The prompt code (Pi) corresponding to the evaluation carried out was included in this analysis (see Annex 2).

4.1 Hetero-intelligence performance testing: contextualisation of overtourism

ChatGPT shows reasonable capabilities regarding contextualising overtourism while presenting the shortcomings summarised in Table 3. This AI is reasonably capable of providing a basic understanding of overtourism (P3). The output acknowledges this tool's training using tourism-related data and present a comprehensive definition of overtourism (P5 and P7), including causes (P10), consequences (P11) and a ranking of affected destinations (P16), such as Santorini and Barcelona. The results generated are validated against the relevant literature (Buitrago et al., 2023; Higgins-Desbiolles, 2018; Joppe, 2018; Mihalic, 2020; Rahmafitria et al. 2020; Santos-Rojo et al., 2023), thereby confirming the output's accuracy.

Table 3- ChatGPT Capabilities and Limitations to Contextualise Overtourism

STAGES	CHATGPT CAPABILITIES	CHATGPT LIMITATIONS
Contextualisation of overtourism	*C1, C2, C4. Capability to carry out a general contextualisation *C7, C8. Self-recognition of some limitations (L2a, L6a, L6d). Accessibility, affordability and availability.	*L1. Limited Depth contextualisation: lack conceptual precision, incomplete or non-structured answers. *L2. Limitations for specificity in concrete contexts (L2a, L2b). *L6, L7. General and technical limitations: L6a. L6b, L6c and L6d (bibliography). Possible ethical and/or legal concerns.

Notes: The same codes for capabilities (C*i*) and limitations (L*i*) have been used as those included in tables 1 and 2.

Source: Own elaboration.

The LLMs thus prove effective in summarising vast amounts of information (P_2) and presenting it in technical language still accessible to non-experts. Without LLMs, acquiring such extensive insights into overtourism would require significant research (P_6). ChatGPT not only defines and describes overtourism but also provides a clear understanding of this problem's impact on sustainable tourism planning (P_6), offers guidelines for policy formulation (P_{14}) and summarises the skills that a good planner should have (P_2). However, limitations are evident especially in the structure and completeness of answers related to the causes (P_{10}) and consequences (P_{11}) of overtourism. The LLM also failed to distinguish between supply, demand, and management factors, and some critical aspects are missing (P10), such as the growing middle class, labour market changes, and infrastructure's increasing privatisation.

Other issues arise when ChatGPT is asked about a specific destination, in this case, Venice (P_1). The model struggles to provide specific causes of overtourism in this city, demonstrating its limitations in understanding particular contexts (P_1). Precision is a problem with technical concepts, such as carrying capacity and references to sources (P_8) – most notably inaccuracies and allusions. For example, ChatGPT responds to prompt P_8 with a reference to a source on IA and overtourism (i.e.: "*Sigala, M. [2019]. Tourism and artificial intelligence: Where are we heading? Journal of Tourism Futures, 5(1), 1–5*"). Although the author and the journal are real and widely recognised in tourism research, the referenced paper is non-existent. This issue appeared in all the recommended bibliography. Allusions or hallucinations are some of ChatGPT's most important limitations already found in other fields (Dwivedi et al., 2023; Mich & Garigliano, 2023). This content could contribute to plagiarism and ethical and legal problems if the information is used without recourse to a human specialist in this field who can adequately screen the answers.

In conclusion, ChatGPT offers a broad understanding of overtourism but shows limitations in its responses' precision and contextualisation, highlighting the need for a collaborative approach involving human expertise to ensure more accurate and reliable results (P_1). This LLM provides a broad understanding of overtourism yet has shortcomings in accuracy and contextualisation. A hetero-intelligence approach is crucial for more accurate findings, especially regarding specific cases and technical concepts.

4.2 Hetero-intelligence performance testing: diagnosing and monitoring sustainable tourism planning regarding overtourism

When analysing and scrutinising overtourism, ChatGPT presents satisfactory capabilities but more significant limitations than previous tasks (see Table 4). This hetero-intelligence performance test provided important confirmation of how human intelligence and AI's interactions can contribute to both retrieving and processing quantitative and qualitative information that can significantly improve tourism planners' initial diagnoses of overtourism's causes and impacts. In particular, human intelligence was revealed as essential to reaching sound conclusions. This finding supports the need for a new conceptualisation of hetero-intelligence as a combination of human intelligence and AI under human supervision when experts deal with complex phenomena in the social sciences.

Table 4: ChatGPT Capabilities and limitations on the diagnosis and monitoring of overtourism

STAGES	CHATGPT CAPABILITIES	CHATGPT LIMITATIONS
Diagnosis and monitoring of overtourism	*C1. Collection of preliminary information for initial diagnosis: interests' contrast, variables, indicators and specific sources *C2. Recovery of some quantitative and qualitative data *C3. Preliminary analysis: it does a general evaluation based on previously identified qualitative and quantitative approaches. *C7, C8. Self-recognition of some limitations (L2a, L4a, L5, L6a, L6d). Accessibility, affordability and availability.	*L1. Limitations for an accurate final diagnosis include a lack of precision in technical concepts, generic, incomplete answers, and errors. *L2. Limitations for a specific final diagnosis: problems in understanding specific contexts and concretising destination information. *L4a. Limitations in generating new primary information (conducting surveys or interviews, designing indicators). *L5. Limitations for critical analysis of multidimensional problems in changing environments and for predicting future scenarios. *L6, L7. Technical limitations: L6a. L6b, L6c and L6d. Ethical and/or legal problems.

Notes: The same codes for capabilities (Ci) and limitations (Li) have been used as those included in tables 1 and 2.

Source: Own elaboration.

This study examined ChatGPT's performance in terms of identifying necessary information for an initial diagnosis of the selected problem. This model successfully retrieves basic data and provides a starting point by gathering general information on potential stakeholders (P15), identifying variables and indicators for further research and tracing existing sources of specific data (P18 and P19). However, the limitations highlighted include that LLMs do not offer all the information needed for precise final diagnoses (P1). Inaccuracies and generic, incomplete information are present in responses regarding academic knowledge about overtourism measures. More specifically, variables used to assess overtourism indicators are inaccurately defined, and answers lack disaggregated spatial and temporal information (P18 and P19). Notably, ChatGPT suggests indicators primarily associated with overtourism consequences, which would make preventing the phenomenon before it occurs challenging. Responses omit both indicators and information sources, such as measures linked to tourism intensity (P10) – a key aspect of overtourism analysis (Buitrago & Yñiguez, 2021; Kirilenko et al., 2023). Limitations are especially prominent in the model's final diagnosis of specific destinations such as Venice. ChatGPT repeats the same information when asked about the measurement of overtourism in general (P18 and P19) and provides insufficient data for Venice (P1, P26 and P27).

On the positive side, ChatGPT contributes to data analyses by providing both quantitative and qualitative results that can be used to assess information gathered for diagnoses (P20). On the negative side, significant shortcomings are present in the specific case of Venice, revealing that the model has a limited analytical capacity when multidimensional problems and non-algorithmic decisions are involved (P1). Another limitation is associated with the dataset with which ChatGPT is trained, which allows it to examine past and present trends but restricts its ability to conduct future-oriented analysis, that is, projections based solely on past data (P1 and P35).

LLMs can identify basic data and provide starting points for diagnosing overtourism. Still, these AI tools exhibit limitations and shortcomings, including inaccuracies and deficits in variables that measure and monitor overtourism. The present results emphasise the importance of hetero-intelligence performance testing of methodologies that combine human intelligence and AI to generate initial diagnoses of complex social science phenomena and the essential role of human intelligence in reaching final conclusions.

4.3 Hetero-intelligence performance testing: stakeholder involvement analysis in overtourism

ChatGPT can help identify and manage stakeholders' involvement as this model is able to identify types of actors and their level of participation. This AI tool, nonetheless, cannot perform various tasks, as shown in Table 5. The academic literature reveals a broad consensus that stakeholder involvement is a key element in sustainable tourism planning (Buitrago et al., 2023; Higgins-Desbiolles, 2018; Martins et al., 2022Mihalic, 2020; Santos-Rojo et al., 2023). ChatGPT's answers also acknowledge this crucial role (P2, P4, P5, P6, P15, P29 and P30). For example, the response to P29 mentions 'the need for tourism planners to balance the interests of various stakeholders, including tourists, local communities, businesses and environmental groups'. ChatGPT can thus be useful for identifying different types of actors and defining their diverse interests and degrees of involvement. These capabilities were highlighted in the answers to prompts P15 and P25 (see Annex 1).

Table 5: ChatGPT Capabilities and Limitations to manage stakeholders' involvement in the context of overtourism

STAGES	CHATGPT CAPABILITIES	CHATGPT LIMITATIONS
Stakeholders involvement	*C1 and C4. Ability to guide planners on the main stakeholders and their different interests. *C6. Ability to make planners aware of the need to involve stakeholders in sustainable planning. *C6. Ability to communicate/make stakeholders aware of existing problems, facilitating their	*L1. Limitations in the depth of stakeholder information include generic, incomplete responses and errors (no information on differences in intra-group interests). *L2. Limitations in the level of specificity of information: L2a and L2b.

	involvement in planning and the search for common ground. *C7 and C8. Self-recognition of some limitations (L2a, L3, L4b, L6a and L6d). Accessibility, affordability and availability	*L3. Lack of social skills. Limitations in capturing and interpreting nuances, emotions or preferences of stakeholders. *L4b. Lack of direct negotiation skills. *L6 and L7. Technical limitations: L6a. L6b, L6c (positive bias favouring sustainable approaches) and L6d. Ethical and/or legal problems.

Notes: The same codes for capabilities (C*i*) and limitations (L*i*) have been used as those included in tables 1 and 2.

Source: Own elaboration.

ChatGPT, however, lacks a capacity for direct negotiation and only offers guidance on how human planners can deal with stakeholder involvement. This model also recognises its limitations in interpreting emotions and the nuances of human conversation, local conditions, and cultural niceties (P_1). ChatGPT has thus failed to develop the social skills necessary for understanding diverse concerns, handling negotiations and grasping local contexts. Generative and conversational AI can analyse patterns in stakeholders' discourses and identify intrinsic and shared ideas, but human intelligence is still essential for working towards a final consensus (Sifat, 2023). ChatGPT provides responses to prompts about overtourism stakeholders and their interests (P_{15} and P_{25}), although these answers reveal limited accuracy, which is important for planners. ChatGPT also provides a comprehensive list of different types of potential stakeholders that falls short of revealing differences in interests. For instance, this model does not acknowledge that a local artisan's concerns may conflict with those of a multinational cruise company or that national and local administrations may have misaligned interests (Goodwin, 2021). Governance challenges in sustainable tourism planning are ignored in ChatGPT's responses. In addition, biased statements appear in this model's answers, such as an assumption that all tourists seek authenticity and cultural immersion (response to P_{15}). The lack of attention paid to context is especially notable in answers related to Venice (P_{15} and P_{25}), in which biased and erroneous statements assert that tourists' interests align with responsible tourism practices and Venetian resources' preservation.

LLMs inadequately analyse stakeholder involvement in overtourism, so their output is considered inadequate to perform the required tasks. As a result, human intelligence is crucial for understanding nuances, addressing governance challenges and discerning biases in AI-generated responses, especially in specific contexts.

4.4 Hetero-intelligence performance testing: policy- and decision-making processes

To present ChatGPT's capabilities and limitations in terms of policy- and decision-*making* processes, this stage of sustainable tourism planning was divided into three substages – definition of measures, decision-making and implementing actions (see Table 6). An evaluation of this model's output revealed that, besides offering general strategies for how to carry out sustainable tourism planning (P_{29}), ChatGPT provides a list of possible measures to apply in destinations struggling with overtourism (P_{30}). In addition, this LLM shares information on good practices that are already being implemented in destinations suffering from this problem (P_{32}). This information can help policy-makers formulate specific policies for the areas they manage. The information summarised in ChatGPT's responses includes key aspects underlined by the literature, but the lists of suggested measures present limitations that must be pointed out to manage them properly.

Table 6: Capabilities and limitations of ChatGPT concerning policy and decision-making

STAGES	CHATGPT CAPABILITIES	CHATGPT LIMITATIONS
Policy- and decision-making		
Definition of measures to carry out	*C1 and C4. Ability to provide guidance on possible measures. *C1. Ability to report examples of good practice. *C5. Ability to stimulate critical and creative thinking of the human planner through brainstorming that may involve interaction with the Chat. *C7, C8. Self-recognition of some limitations (L2a, L3c, L6a and L6d). Accessibility, affordability and availability	*L1. Limitations in the measures' depth level: lack of precision in technical concepts, generic, unstructured, incomplete answers and errors. *L2. Limitations in proposing measures for specific destinations: lack of concreteness, errors in generalisation. *L3c. Limitations in creative thinking (proposing measures that involve different approaches to existing ones). *L6, L7. Technical constraints: L6a. L6b, L6c, L6d. Ethical and/or legal issues.

Decision-making	*C1, C4. Ability to provide guidance on benefits and risks of measures. *C3b, c. Ability to report suggestions and opinions on actions and to identify trends *C7, C8. Self-recognition of some limitations (L2a, L3a, L3b, L5, L6a, L6d). Accessibility, affordability, and availability	*L2a. Limitations in interpreting specific contexts. *L3. L3a. Limitations in capturing and interpreting nuances, emotions or preferences of users. Lack of personal experience L3b. Limitations in critical thinking. L3d. Limitations in adapting to changing circumstances. *L5. Limitations for in-depth analysis and predicting future scenarios. *L6 and L7. Technical limitations: L6a. L6b, L6c, L6d. Ethical and/or legal issues.
Implementation of actions	*C6. Support tool for the implementation of some measures (e.g. awareness-raising/information to tourists or residents)	*L4c. Due to its characteristics, it does not have the capacity to directly implement the measures decided.

Notes: The same codes for capabilities (Ci) and limitations (Li) have been used as those included in tables 1 and 2.

Source: Own elaboration.

The present study compared ChatGPT's contributions based on the literature (Butler & Dodds, 2022; Goodwin, 2021; Higgins-Desbiolles, 2018; Rahmafitria et al., 2020; Wall, 2020), which again reveals a lack of academic-technical precision in some answers. The list of proposed measures is randomly structured without basis on any criteria (i.e. objective pursued and type of instrument). The measures included are also not exhaustive, and they focus more on the objectives pursued than on the specific strategies to be implemented. In addition, significant omissions can be detected regarding measures already applied in many destinations. The response to prompt P30 omits solutions such as demarketing, destination marketing organisation, destination management platforms, and priority passes for residents.

In alignment with Iskender's (2023) results, clear evidence exists that ChatGPT, which is trained with historical information, has shortcomings in measures involving innovative or novel thinking, especially in new situations. When prompted to recommend innovative measures (P31), this model expands the initial list by including information related to advanced technologies, such as destination management platforms. These measures comprise technological innovations, but they fail to provide viewpoints or approaches different from those already existing in the databases used to train ChatGPT. Hetero-intelligence conversations can be valuable in terms of brainstorming, stimulating critical and creative thinking and allowing human intelligence to develop new approaches or ideas (P4). This is another example of how hetero-intelligence, as understood in this research, is crucial to achieving sustainable tourism planning, which reinforces Dwivedi et al.'s (2023) findings.

ChatGPT acknowledges that 'there is no one-size-fits-all solution to overtourism' (response to P29), yet its answer to the prompt about concrete measures for managing overtourism in Venice (P36) is nearly identical to the response when the destination is unspecified (P30). This broad generalisation even leads to notable mistakes, such as recommending the use of bicycles in this city of canals (P36). To validate this limitation, ChatGPT was also asked to list measures to alleviate overtourism in other destinations, specifically two identified in the ranking of destinations suffering from overtourism: Barcelona (P38) and Santorini (P39). ChatGPT's potential application becomes even more restricted when this model is asked to make decisions about concrete measures for specific contexts. AI can guide planners on the potential risks and benefits of various measures proposed for Venice (P37), but the reported answers are quite generalised. Critical analysis, which is essential for decision-making, has been identified as a primary shortcoming of ChatGPT (response to P1) (Dwivedi et al., 2023). To assess this limitation's implication, the present study's hetero-intelligence conversation included opinion questions related to how relevant different measures are and what ethical issues could arise while applying different strategies to address overtourism in general (P33 and P34) and in a specific case such as Venice (P42 and P43). All the model's answers acknowledge ChatGPT's limitations, thereby emphasising the need for human intelligence in decision-making (e.g. 'Human judgment and critical thinking are crucial in interpreting and contextualising the information provided by ChatGPT to ensure effective and responsible decision-making in managing overtourism' [response to P4]).

Another self-acknowledged shortcoming is ChatGPT's difficulty in predicting future scenarios for decision-making (response to P1). To test this limitation, a question about future challenges related to overtourism was included (P35). An evaluation of the response found that this model made a general prediction, namely, 'overtourism is a challenge that will likely persist and may even grow in the future' (response to P35), based on challenges mentioned in the academic literature (Buitrago et al., 2023; Butler & Dodds, 2022; Santos-Rojo et al., 2023).

ChatGPT provides useful information for sustainable tourism planning, but its answers have limitations in precision, structure and exhaustiveness when proposing measures to deal with overtourism. The above hetero-intelligence performance testing highlighted the potential for more efficient and effective policy- and decision-making processes through human intelligence and AI cooperation in sustainable tourism planning. This approach provides a way to overcome ChatGPT's identified shortcomings.

5. Conclusions and implications

By systematically developing and testing an exploratory hetero-intelligence methodology, this research sought to demonstrate how the co-performance of human intelligence and AI can address complex tourism issues in a more effective and reliable way and offer significant advantages for effective and sustainable tourism policy-making. It enhances decision-making by combining AI's data processing capabilities with human contextual understanding and supervision. It ensures comprehensive analysis through a bi-dimensional scale that addresses both general trends and localised issues. The seven-step process of the developed framework guarantees a rigorous and systematic approach. By identifying AI's capabilities and limitations, the methodology effectively integrates AI into planning while highlighting areas requiring human expertise. Iterative improvements through peer review ensure adaptability and evolution with new insights.

The results reveal that a hetero-intelligence approach can be useful within sustainable tourism planning once the capabilities and limitations of both human intelligence and AI are properly identified. The results also confirmed that the co-performance of AI and human intelligence needs to use a methodological framework based on empirically rigorous steps that can provide unbiased outcomes.

5.1Theoretical Implications

The proposed methodology constitutes a novelty in tourism literature and makes an epistemological contribution to this field. Specific guidelines are provided for how to apply the hetero-intelligence methodology so that it can be replicated in other sustainable tourism planning contexts.

The findings highlight the importance of a hetero-intelligence framework to be applied under human supervision, especially in the social sciences, where contextual understanding and critical thinking are needed. This approach aligns with emerging theories on the collaborative potential between humans and AI, suggesting that while AI can handle vast data synthesis and initial diagnostics, human expertise is indispensable for nuanced analysis and final decision-making.

5.2 Practical Implications

The results have varied practical implications, including the need for human tourism planners to strengthen some of their skills to take full advantage of AI's contributions. For tourism planners, the integration of AI can significantly speed up initial data gathering and provide a foundational understanding of complex issues. However, the limitations in AI's contextual accuracy and depth mean that human experts must play a critical role in validating and refining AI-generated insights. For local communities and businesses, AI tools can identify diverse stakeholder interests but may fall short in addressing nuanced and conflicting interests, underscoring the need for human-mediated negotiation and decision-making processes. When using LLMs in planning, tourism professionals and policy-makers should have extensive knowledge of tourism and destination planning and capabilities in prompt engineering to guide conversations between these models and human experts, ask the right questions and interpret the answers correctly. Otherwise, the hetero-intelligence approach's potential will not be fully achieved. The study's findings stress the importance of social skills to ensure planners grasp stakeholders' different sensitivities in negotiations, understand contexts' nuances and have the required cognitive skills to engage in critical and creative thinking. In addition, The framework and its application, as exemplified in this study, may be useful in guiding tourism planners and policy-makers in the process of taking full advantage of AI tools.

5.3 Limitations and suggestions for future studies

One of LLMs' main limitations is that they are evolving continuously, which implies the need to monitor their output periodically. This research included a series of checks of the output produced in conversations with ChatGPT-4 between December 2022 and October 2023, although no substantial changes were found. Another shortcoming may be that this experiment was conducted only with ChatGPT-4 and was not replicated with other LLMs.

Suggested future studies could involve conducting more empirical research to test and validate the proposed methodology. Additional studies can use hetero-intelligence to assess other sustainability challenges in tourism. A further important avenue would be to monitor the evolution of LLMs' contribution to hetero-intelligence conversations in the near future. Also, there is a lack of studies addressing ethical and legal issues related to the use of LLMs, particularly in sustainable tourism planning.

Acknowledgements

This research is supported by the Research Groups of Research in Applied Economics (SEJ132, SEJ 258. Plan Andaluz de Investigación, Junta de Andalucía, Spain). This paper is also financed by National Funds provided by FCT—Foundation for Science and Technology through project UIDB/04020/2020 with https://doi.org/10.54499/UIDB/04020/2020.

Credit author statement

All authors have contributed equally. All authors have read and agreed to the published version of the manuscript.

Declaration of competing interest: None

References

Adeoye-Olatunde, O.A., & Olenik, N.L. (2021). Research and scholarly methods: Semi-structured interviews. *Journal of the American College of Clinical Pharmacy*, 4(10), 1358-1367. https://doi.org/10.1002/jac5.1441

Alberts, I., Mercolli, L., Pyka, T., Prenosil, G., Shi, K., Rominger, A., & Afshar-Oromieh. A. (2023). Large language models (LLM) and ChatGPT: what will the impact on nuclear medicine be?. *European Journal of Nuclear Medicine and Molecular Imaging*, 50, 1549–1552. https://doi.org/10.1007/s00259-023-06172-w

Alipour, H., Pendar, N., & Roy, K. (024). ChatGPT alternative solutions: Large Language Models survey. *Arxiv*, 2403.14469v1. https://doi.org/10.48550/arXiv.2403.14469

Biswas, S. (2023). ChatGPT and the future of medical writing. *Radiology*, 307(2), e223312. https://doi.org/10.1148/radiol.223312

Breakspear, A. (2013). A new definition of intelligence. *Intelligence and National Security*, 28(5), 678-693.

Buitrago, E.M., & Yñiguez, R. (2021). Measuring overtourism: A necessary tool for landscape planning. *Land*, 10(9), 889. https://doi.org/10.3390/land10090889

Buitrago, E.M., Foronda-Robles, C. & Yñiguez-Ovando, R. (2023). A literature review on overtourism to guide the transition to responsible tourism. *Revista de Estudios Andaluces*, 45, 71-90. https://dx.doi.org/10.12795/rea.2023.i45.04

Buitrago-Esquinas, E.M., Puig-Cabrera, M., Santos, J.A.C., Custódio-Santos, M., & Yñiguez-Ovando, R. (2024). Developing a hetero-intelligence methodological framework for sustainable policy-making based on the assessment of large language models. *MethodsX*, 12, 102707. https://doi.org/10.1016/j.mex.2024.102707

Busch, K., Rochlitzer, A., Sola, D., & Leopold, H. (2023). Just tell me: Prompt engineering in business process management. In *International Conference on Business Process Modeling, Development and Support* (pp. 3-11). Cham: Springer Nature Switzerland.

Butler, R.W., & Dodds, R. (2022). Overcoming overtourism: a review of failure. *Tourism Review*, 77(1), 35-53. https://doi.org/10.1108/tr-04-2021-0215

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

Celata, F., & Romano, A. (2022). Overtourism and online short-term rental platforms in Italian cities. *Journal of Sustainable Tourism*, 30(5), 1020-1039. https://doi.org/10.1080/09669582.2020.1788568

Çelik, M. N., & Çevirgen, A. (2021). The role of accommodation enterprises in the development of sustainable tourism. *Journal of Tourism and Services*, 12(23), 181–198. https://doi.org/10.29036/jots.v12i23.264

Crabolu, G., Font, X., & Eker, S. (2023). Evaluating policy complexity with Causal Loop Diagrams. *Annals of Tourism Research*, 100, 103572. https://doi.org/10.1016/j.annals.2023.103572

Dellermann, D., Calma, A., Lipush, N., Weber, T., Weigel, S., & Ebel, P. (2019). The future of human-ai collaboration: a taxonomy of design knowledge for hybrid intelligence systems. Proceedings of the 52nd Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2019.034

Dwivedi, Y.K., Kshetri, N., Hughes, L., Slade, E.L., Jeyaraj, A., Kar, A.K., ... & Wright, R. (2023). So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

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

Eggmann, F., Weiger, R., Zitzmann, N.U., & Blatz, M.B. (2023). Implications of large language models such as ChatGPT for dental medicine. *Journal of Esthetic and Restorative Dentistry*. https://doi.org/10.1111/jerd.13046

Flick, U. (2023). *An introduction to qualitative research* (7th ed.). London: Sage

Goodwin, H. (2021). City destinations, overtourism and governance. *International Journal of Tourism Cities*, 7(4), 916-921. https://doi.org/10.1108/IJTC-02-2021-0024

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. https://doi.org/10.1080/19368623.2023.2211993

Henriques, H.J.G., Almeida, C.R., & Ramos, C.M.Q. (2024). The Application of artificial intelligence in the tourism industry: a systematic literature review based on Prisma methodology. *Journal of Tourism, Sustainability and Well-Being*, 12(1), 65-86. https://doi.org/10.34623/hkqk-ht95

Higgins-Desbiolles, F. (2018). Sustainable tourism: Sustaining tourism or something more? *Tourism Management Perspectives*, 25, 157-160. https://doi.org/10.1016/j.tmp.2017.11.017

Iskender, A. (2023). Holy or unholy? Interview with open AI's ChatGPT. *European Journal of Tourism Research*, 34, 1-12. https://doi.org/10.54055/ejtr.v34i.3169

Ivanov, S., & Soliman, M. (2023). Game of algorithms: ChatGPT implications for the future of tourism education and research. *Journal of Tourism Futures*, 9(2), 214-221. https://doi.org/10.1108/JTF-02-2023-0038

Izzo, F., & Picone, Q. (2022). Defining an integrated and computed methodology approach for sentiment and psychographic analysis in tourism research. *Journal of Tourism and Services*, 13(25), 1–21. https://doi.org/10.29036/jots.v13i25.393

Joppe, M (2018). Tourism policy and governance: Quo vadis?, *Tourism Management Perspectives*, 25, 201-204. https://doi.org/10.1016/j.tmp.2017.11.011

Kirilenko, A.P., Ma, S., Stepchenkova, S.O., Su, L., & Waddell, T.F. (2023). Detecting early signs of overtourism: bringing together indicators of tourism development with data fusion. *Journal of Travel Research*, 62(2), 382-398. https://doi.org/10.25384/SAGE.c.5771173.v1

Lacárcel, F.J.S. (2022). Main uses of artificial intelligence in digital marketing strategies linked to tourism. *Journal of Tourism, Sustainability and Well-Being*, 10(3), 15-226. https://doi.org/10.34623/mppf-r253

Lo, L.S. (2023). The CLEAR path: A framework for enhancing information literacy through prompt engineering. *The Journal of Academic Librarianship*, 49(4), 102720. https://doi.org/10.1016/j.acalib.2023.102720

Martins, P. G., Ferreira, A. M. A., & Costa, C. M. (2022). Tourism and third sector organizations: synergies for responsible tourism development?. *Tourism & Management Studies*, 18(1), 7-16. https://doi.org/10.18089/tms.2022.1801

Majid, G.M., Tussyadiah, I., Kim, Y.R., & Pal, A. (2023). Intelligent automation for sustainable tourism: a systematic review. *Journal of Sustainable Tourism*, 31(11), 2421-2440.

Mich, L., & Garigliano, R. (2023). ChatGPT for e-Tourism: a technological perspective. *Information Technology & Tourism*, 25, 1-12. https://doi.org/10.1007/s40558-023-00248-x

Mihalic, T. (2020). Concpetualising overtourism: A sustainability approach. *Annals of Tourism Research*, 84, 103025. https://doi.org/10.1016/j.annals.2020.103025

OpenAI (2023). GPT-4 is OpenAI's most advanced system, producing safer and more useful responses. Available from: https://openai.com/product/gpt-4

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

Rahmafitria, F., Pearce, P.L., Oktadiana, H. & Putro, H.P.H (2020). Tourism planning and planning theory: Historical roots and contemporary alignment. *Tourism Management Perspectives*, 35, 100703. https://doi.org/10.1016/j.tmp.2020.100703

Rodríguez, P., Valencia-Arias, A., Garcés-Giraldo, L., Castañeda, L., Moreno, G., & Benjumea-Arias, M. (2023). Tendencias en el uso de inteligencia artificial en el sector del turismo. *Journal of Tourism & Development*, 40, 81-92. https://doi.org/10.34624/rtd.v40i0.31447

Santos-Rojo, C., Llopis-Amorós, M., & García-García, J.M. (2023). Overtourism and sustainability: A bibliometric study (2018–2021). *Technological Forecasting and Social Change*, 188, 122285. https://doi.org/10.1016/j.techfore.2022.122285

Sifat, R.I. (2023). ChatGPT and the future of health policy analysis: potential and pitfalls of using ChatGPT in policy-making. *Annals of Biomedical Engineering*, 51, 1357-1359. https://doi.org/10.1007/s10439-023-03204-2

Skavronskaya, L., Hadinejad, A., & Cotterell, D. (2023). Reversing the threat of artificial intelligence to opportunity: a discussion of ChatGPT in tourism education. *Journal of Teaching in Travel & Tourism*, 23(2), 253-258. https://doi.org/10.1080/15313220.2023.2196658

Torres, R.M. (2021). The empty boxes of Venice: Overtourism—Conflicts, politicisation and activism. In *The power of new urban tourism* (pp. 147-160). Routledge.

Ülkü, A. (2023). Artificial intelligence-based large language models and integrity of exams and assignments in higher education: the case of tourism courses. *Tourism & Management Studies*, 19(4), 21-34. https://doi.org/10.18089/tms.2023.190402

Valentinas, N., Petrokė, I., Urbański, M., & Soboleva, O. (2022). Macroeconomic factors influencing the development of the sharing economy in the Lithuanian tourism sector. *Journal of Tourism and Services*, 13(25), 248–266. https://doi.org/10.29036/jots.v13i25.437

Wall, G. (2020). From carrying capacity to overtourism: a perspective article. *Tourism Review*, 75(1), 212-215. https://doi. org/10.1108/TR-08-2019-0356

Wong, I.A., Lian, Q.L., & Sun, D. (2023). Autonomous travel decision-making: An early glimpse into ChatGPT and generative AI. *Journal of Hospitality and Tourism Management*, 56, 253-263. https://doi.org/10.1016/j.jhtm.2023.06.022

Yrigoy, I., Horrach, P., Escudero, L., & Mulet, C. (2024). Co-opting overtourism: tourism stakeholders' use of the perceptions of overtourism in their power struggles. *Journal of Sustainable Tourism*, 32(4), 818–834. https://doi.org/10.1080/09669582.2023.2178445

Zhang, Y., & Prebensen, N.K. (2024). Co-creating with ChatGPT for tourism marketing materials. *Annals of Tourism Research Empirical Insights*, 5(1), 100124. https://doi.org/10.1016/j.annale.2024.100124

ANNEX 1. Questionnaire to guide the ChatGPT interview on overtourism

Part 1: Self-recognition of ChatGPT as a sustainable tourism planner

P_1. What are your own limitations for sustainable tourism planning?
P_2. What are the main abilities and skills that should have a good tourism planner? Do you have them?
P_3. Does your knowledge not meet these required abilities and skills?
P_4. How do you think that ChatGPT could affect overtourism?

Part 2: Contextualisation and diagnosis of overtourism

2a. General overtourism

P_5. What is overtourism?
P_6. How does overtourism affect sustainable tourism planning in a destination?
P_7. What are the concepts of overtourism, gentrification, turistification, turismophobia, and Trexit?
P_8. Where did you take these definitions from?
P_9. Could you give me references that combine AI and overtourism?
P_{10}. What are the causes of overtourism?
P_{11}. What are the consequences of overtourism?
P_{12}. How does overtourism affect tourists?
P_{13}. How does overtourism affect residents?
P_{14}. Does overtourism affect social, economic, and environmental sustainability?
P_{15}. What are the main stakeholders typologies involved in overtourism and what are their main interests in tourism planning?
P_{16}. What is the ranking of top destinations suffering from overtourism?
P_{17}. What criteria did you use to make this ranking?
P_{18}. What kind of information should we collect to make a complete diagnosis of overtourism in a destination?
P_{19}. Concerning the overtourism symptoms, could you give me some quantitative approaches so that a destination can be diagnosed?
P_{20}. You did not give me reference values, could you please include them in your explanation?
P_{21}. Considering the following information, could we consider that there is overtourism in a certain destination?: 10 visitors per squared meters transportation over 90% rental spending bigger than 30% of income of residents a proportion of 3:1 visitors per resident more than 50% dependence on tourism

2b. Venice

P_{22}. Is Venice (Italy) a destination that suffers from overtourism? Why?
P_{23}. What are the main causes of overtourism in Venice?
P_{24}. What are the main consequences of overtourism in Venice? Please be specific
P_{25}. What are the main stakeholders involved in overtourism and what are the main what are their interests in tourism planning in Venice?
P_{26}. Could you give me quantitative data on overtourism in Venice?
P_{27}. What is the most updated data on the following indicators?
P_{28}. Venezia has the following situation, should we consider that there is overtourism in this city according to this information? (Please check if the information is correct): 5 visitors per squared meters transportation over 60% rental spending bigger than 20% of the income of residents a proportion of 1:1 visitors per resident more than 30% dependence on tourism

Part 3: Policy formulation to manage overtourism and decision-making

3a. General overtourism

P_{29}. What are the main guidelines for policy-making to assure an effective management of overtourism?
P_{30}. How can overtourism be effectively managed in a destination?
P_{31}. Could you give me any innovative solution to manage overtourism?
P_{32}. Could you give me some good practices that were already designed to manage overtourism? Please indicate the location and the destination in which these practices were implemented.
P_{33}. Do you think that tourism should be limited? What is your opinion on overtourism concerning its main consequences?
P_{34}. Do you think it is ethical to deprive residents of a city center because of tourism?
P_{35}. What are the future challenges concerning overtourism?

3b. Venice

P_{36}. Could you make a list of measures that could contribute to alleviating overtourism in Venice?
P_{37}. What are the main risks and benefits if we apply these measures in Venice? Please be specific.
P_{38}. Could you make a list of measures that could contribute to alleviating overtourism in Barcelona?
P_{39}. Could you make a list of measures that could contribute to alleviating overtourism in Santorini?
P_{40}. What measures were already developed to address overtourism in Venice?
P_{41}. How much is the tourist tax in Venice?
P_{42}. Should Venice ban tourism?
P_{43}. Should Venice keep the tourist tax or not? It is a big dilemma among its stakeholders.

ANNEX 2. ChatGPT capabilities & limitations for sustainable tourism planning: Assessing a conversation on overtourism

Prompts[(a)]	Capabilities[(b)]								Limitations[(c)]						
	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	L_1	L_2	L_3	L_4	L_5	L_6	L_7
P_1	x			x		x	x	x	b,c					x	X
P_2	x			x		x	x	x	b,c					x	X
P_3	b		x	x		x	x	x			x	x	x	x	X
P_4	b		a		x	x	x	x	b,c					x	X
P_5	x					x		x	a,b				x	x	X
P_6	x		a,b			x		x	a,b,c		b,c		x	x	X
P_7	x					x		x	a,b,c				x	x	X
P_8				x			x	x						x	X
P_9								x	d					x	X
P_{10}	x		a		x	x		x	a,b,c		b,c		x	x	X
P_{11}	x		a		x	x		x	a,b,c		b,c		x	x	X
P_{12}	x		a		x	x		x	a,b,c		b,c		x	x	X
P_{13}	x		a		x	x		x	a,b,c		b,c		x	x	X
P_{14}	x		a		x	x		x	a,b,c		b,c		x	x	X
P_{15}	x		a		x	x		x	x		x	b	x	x	X
P_{16}	x		a,b	x	x		x	x	b,c	a,b			x	x	X
P_{17}	b,c		a,b	x	a		x	x	b,c	b			x	x	X
P_{18}	x		b					x	x	a		a	x	x	X
P_{19}	x		b					x	x	a		a	x	x	X
P_{20}	x	x	b	x	a	x	x	x	x	a	b		x	x	X
P_{21}	b		a	x	a		x	x	b,c	b	b,d		x	x	X
P_{22}	x	x	a,b			x		x	a,b,c	a,b	b		x	x	X
P_{23}	x		a		x	x		x	a,b,c	a,b	b		x	x	X
P_{24}	x	x	a		x	x		x	a,b,c	a,b	b		x	x	X
P_{25}	x		a,b		x	x		x	x	x	x	b	x	x	X
P_{26}	b				x	x	x	x	b,c	a				x	X
P_{27}	b						x	x	x	x		a		x	X
P_{28}	b,c		a,b		x		x	x	x	x	x		x	x	X
P_{29}	x		a,b		x	x		x	b,c				x	x	X
P_{30}	x		b		x	x		x	x		x		x	x	X
P_{31}	x		x	x	x	x		x	b		b,c		x	x	X
P_{32}	x	x		x	x			x	b,c	a			x	x	X
P_{33}	x		a,b	x	x	x	x	x	b					x	X
P_{34}	x		a,b	x	x	x		x	b					x	X
P_{35}	x		x	x	x	x		x	b,c		b,c		x	x	X
P_{36}	x				x	x		x	x	x	x		x	x	X
P_{37}	x		x		x	x		x	b,c	a,b	x		x	x	X
P_{38}	x				x	x		x	x	x	x		x	x	X
P_{39}	x				x	x		x	x	x	x		x	x	X
P_{40}	x	x		x	x	x		x	x	a,b				x	X
P_{41}	x	x		x			x	x	a,b,c			c	x	x	X
P_{42}	b,c		x	x	x	x		x	a,b,c	a,b	x		x	x	X
P_{43}	b,c		a	x	x	x		x	a,b,c	a,b	x	c	x	x	X

Notes: [(a)] Prompts (P_i) that guided the conversation are shown in Annex 1. [(b), (c)] The list of identified Capabilities (C_i) and Limitations (L_i) are shown in Tables 1-2. Each "x" represents that the capability or limitation was identified in the corresponding prompt; "a", "b", "c" or "d" is given in replace of "x" when only this element is identified from the capability/limitation on that prompt.

Source: Own elaboration.