



# Market research and knowledge using Generative AI: the power of Large Language Models

Macarena Estevez<sup>a</sup> , María Teresa Ballestar<sup>a,b,\*</sup> , Jorge Sainz<sup>a,c</sup> 

<sup>a</sup> Universidad Rey Juan Carlos, Madrid, Spain

<sup>b</sup> Department of Social Sciences, Lappeenranta-Lahti University of Technology (LUT University), Finland

<sup>c</sup> IPR; University of Bath, Bath, UK

## ARTICLE INFO

### JEL classification:

C83

M21

L66

### Keywords:

Generative AI

Marketing

Market research

LLMs

## ABSTRACT

Generative artificial intelligence (GAI) is rapidly transforming the marketing industry with its capabilities in data analysis, personalisation, strategic optimisation, and content generation, providing powerful tools for businesses aiming to establish or maintain a strong brand position. This research examines the application of Large Language Models (LLMs) for market research, enabling marketers to create insights that capture the complexities of market behaviour and customer psychology, helping them to connect with their target audiences more meaningfully and effectively. Particularly, it aims to evaluate the extent to which GAI can replicate traditional market research. We conducted a comprehensive survey on beer consumption in Spain, covering all conversion funnel stages, from Brand Awareness to Purchase. The study was replicated using four prominent LLMs: ChatGPT (OpenAI), Gemini (Google), Claude (Anthropic) and LLaMa (Meta), and the results of these LLMs were compared with those of the traditional survey using a collection of statistical methods. Our results show that LLMs are valuable for market research, offering significant insights as reliable proxies. This represents a competitive advantage by making studies of this kind more accessible and cost-effective, benefitting companies of all sizes. However, LLMs cannot fully replicate traditional methods and present result variability, introducing risks in decision-making because of potential errors in data generation that are complex to estimate without benchmarking.

## Introduction

Artificial intelligence (AI) is rapidly transforming the marketing landscape, providing powerful tools for companies to consolidate and maintain a strong brand position. One of the most important contributors to this transformation is the rise of Generative AI (GAI), a subset of machine learning that uses algorithms to create new data, such as text, images, and code, based on existing data, driven by advanced Large Language Models (LLMs) (Kshetri et al., 2023). This technology promises unprecedented scale and speed in gaining market insights. For example, LLMs can process and analyse millions of online reviews in real time to identify emerging consumer preferences, which would take human analysts weeks or months (Cano-Marín, 2024; Yenduri et al., 2024).

The global LLM market is experiencing significant growth. Projections show that it will increase from USD 6.4 billion in 2024 to USD 36.1 billion in 2030, at an impressive compound annual growth rate of

33.2 % (MarketsandMarkets, 2024). This growth is in response to the increasing demand for advanced natural language processing capabilities in various industries, including marketing and consumer research (Pagani & Wind, 2024). This article focuses on the latter and explores how LLMs are revolutionising brand positioning by enabling marketers to connect meaningfully and effectively with their audiences. It discusses the application of LLMs in competitive market analysis and compares their capabilities to traditional research methods.

For marketers, particularly those in small and medium-sized enterprises (SMEs), LLMs offer an unprecedented opportunity to conduct sophisticated market research and customer segmentation without the traditionally high costs associated with large-scale surveys or focus groups (González-Padilla et al., 2025; Zhu & Zhang, 2023). Regarding the automation of tasks such as data collection, analysis or message personalisation, LLMs reduce overhead and execution times, enabling marketing strategies to be deployed more quickly and at scale. Moreover, as these models can process and synthesise vast amounts of

\* Corresponding author.

E-mail address: [teresa.ballestar@urjc.es](mailto:teresa.ballestar@urjc.es) (M.T. Ballestar).

<https://doi.org/10.1016/j.jik.2025.100796>

Received 6 April 2025; Accepted 19 August 2025

Available online 2 September 2025

2444-569X/© 2025 The Author(s). Published by Elsevier España, S.L.U. on behalf of Journal of Innovation & Knowledge. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

unstructured consumer data in real time, they allow for a faster adaptation to market trends and audience preferences, levelling the playing field between SMEs and large corporations (Dwivedi, 2025).

AI is a set of programmes, algorithms, systems, and machines that exhibit aspects of human intelligence, enabling marketers to leverage vast volumes of consumer data (Arora et al., 2024; Park & Ahn, 2024). This dataset includes everything from sentiment analysis on social networks (Nicolescu & Rîpă, 2024; Rostamzadeh et al., 2024; Zhou et al., 2023) to shopping behaviour (Hermann & Puntoni, 2024; Peres et al., 2023) and online behaviour (J. S. Park et al., 2022; Saura & Ruizxisiene, 2025), generating actionable information that facilitates precise personalisation (Campbell et al., 2023; Prados-Castillo et al., 2024; Tafesse & Wien, 2024), and creating highly segmented messages across multiple channels (Cillo & Rubera, 2024; Kumar et al., 2024; Lee et al., 2024).

Emerging literature suggests that the most effective approach is a hybrid model, where GAI complements and enhances traditional methods rather than replacing them entirely (Chiarello et al., 2024; Estevez et al., 2024; Grewal et al., 2024a, 2024b; Park & Ahn, 2024). The main advantages of this integration include cost reduction by automating tasks such as data collection, transcription, and analysis, the ability to process massive volumes of information more quickly and accurately, and the creation of personalised messages tailored to specific market segments, thus optimising marketing strategies (Brand et al., 2024; Goretti et al., 2024; Li et al., 2024).

This article contributes to the debate on the future of market research in the age of AI by examining the extent to which LLMs can complement or replace traditional methods in assessing brand perception and consumer psychology. It addresses key issues such as identifying brand personality traits, mirroring consumer survey responses, generating persuasive personalised messages, using synthetic data, collaborating on market research, and conducting perceptual analysis (Atari et al., 2023; Matz et al., 2024; Wach et al., 2023).

However, current LLMs face significant challenges in replicating the dynamic decision-making processes characteristic of human behaviour in competitive environments. Recent experimental studies have shown that LLMs cannot achieve market equilibrium in simulated scenarios, underscoring the need for additional advances to fully capture the complexities of market behaviour and consumer psychology (Bolukbasi et al., 2016; Brown et al., 2020; Cao & Kosinski, 2024; Demsky et al., 2020; Pagani & Wind, 2024).

Our research compares leading LLMs to traditional market research (TMR) methods, aiming to provide a comprehensive view of AI's potential and limitations in marketing. By exploring the synergies between AI-driven approaches and conventional methods, we seek to contribute to developing more effective and efficient marketing strategies in a constantly evolving digital environment.

This transformation in marketing, driven by GAI and LLMs, demands a reassessment of traditional methodologies and a re-evaluation of how brands understand and engage consumers. In this context, our study contributes to the debate on the future of market research by empirically examining the role of LLMs as both a complement and a potential substitute for traditional survey methods (Juárez-Madrid & Monreal, 2024). The rest of the article is structured as follows. Section 2 presents the review of the literature and theoretical framework. Section 3 describes the data collection, study design, and execution. Section 4 details the empirical analysis stage, including descriptions of the statistical methods used and the presentation of the results along with their explanations. Section 5 summarises the conclusions of this research and offers proposals for future research.

## Theoretical framework

The advent of GAI has sparked growing interest in its application to market research, promising unprecedented scale and speed in obtaining insights (Kshetri, 2023). This ability to analyse vast data sets and

generate real-time knowledge offers companies a potential competitive advantage. However, it is imperative to recognise the inherent limitations of GAI. The quality and potential bias of training data directly influence the accuracy and reliability of the responses generated, with the risk of producing inaccurate or even misleading information (Abumalloh et al., 2024; Campbell et al., 2023; Eloundou et al., 2023; Pagani & Wind, 2024).

Emerging literature suggests that the most effective strategy is a hybrid approach, where GAI complements and enhances traditional methods rather than replacing them entirely (Noy & Zhang, 2023; Peres et al., 2023). Among the most prominent advantages of GAI in this context are 1) cost reduction by automating tasks such as data collection, transcription, and analysis, generating significant savings in time and resources (Noy & Zhang, 2023), 2) scalability, enabling a faster and more accurate analysis of massive volumes of data compared to traditional methods and encompassing a wider range of sources such as social networks, product reviews, and online forums (Kshetri et al., 2024; Peres et al., 2023), and 3) personalisation, facilitating the creation of messages and content tailored to specific consumer segments, thus optimising marketing strategies based on the needs and preferences of the target audience (Hermann & Puntoni, 2024; Kumar et al., 2024; Lee et al., 2024; Wahid et al., 2023). This theoretical framework aims to explore in depth these advantages and limitations and emerging methodologies that combine GAI with traditional approaches to delineate the future of marketing research in the era of AI.

## LLMs

LLMs are a subcategory of foundational AI models that have emerged as a disruptive technology because of their ability to process and generate natural language text (Brown et al., 2020; Devlin, 2018). A language model (LM) is a probabilistic representation of a language, which estimates the probability of occurrence of a given linguistic sequence, be it a word, phrase, or sentence, within that language.

LLMs, similar to conventional LMs, are trained using a self-supervised learning paradigm, where the goal is to predict a hidden word within a textual sequence (Mikolov, 2013). The term 'large' refers to the massive scale of data with which they are trained, often spanning multiple languages and heterogeneous textual corpora, as well as the sheer number of parameters that these models possess. Following Achiam et al. (2023) and Raffel et al. (2020), we define an LLM as a conditional probability distribution  $p(x_n | x_1, \dots, x_{n-1})$  over tokens, where each  $x_i$  belongs to a finite  $V$  vocabulary. In their training, they use a mixture of datasets that include multilingual corpora such as Common Crawl, Wikipedia, and specialised domain data. To handle multiple languages, they incorporate tokenisers and embeddings designed for language-specific characteristics, ensuring semantic consistency across linguistic boundaries.

The base of the text generative process in an LLM is an iterative sampling of the learned distribution to select the next most likely token. At each iterative step, the model evaluates a probability distribution over the vocabulary  $V$ , which indicates the probability that each token is the following  $x_i$  in the sequence, assuming that the model was reading a pre-existing text. Initiating this generative process requires an initial 'conditioning', that is, a set of input tokens  $x_1, \dots, x_{n-1}$ , known as a 'prompt'.

The prompt acts as an initial context that influences the probability distribution over the vocabulary, favouring the selection of tokens semantically and syntactically consistent with the given context. For example, given the prompt 'This is a review...', the model would assign a higher probability to tokens such as 'article' or 'book' compared to tokens such as 'bus' or 'mountain'. Using a distribution function, such as the softmax function, the model stochastically selects a token from a set of plausible candidates. The new token  $x_i$  is incorporated into the sequence, resulting in a new iteration of the process.

Prominent examples of LLMs include models such as BERT

(Bidirectional Encoder Representations from Transformers) (Devlin et al., 2018), T5 (Text-to-Text Transfer Transformer) (Raffel et al., 2020), and the GPT series, including ChatGPT-4 (OpenAI, 2023).

### LLMs and human personality

The intersection between LLMs and human personality is an emerging area of research with significant implications for understanding human behaviour and developing more sophisticated human-computer interaction systems. Although LLMs, as computational models, inherently lack personality, they have demonstrated the ability to capture and reflect human perceptions of third-party personality (Cao & Kosinski, 2024; Kirk & Givi, 2025; Lee et al., 2024). For instance, Kirk and Givi (2025) showed that GPT-4 could infer Big Five personality traits from social media bios with 80 % accuracy relative to human raters.

A key finding in this domain is the capacity of LLMs to predict perceived personality. Recent studies using models such as GPT-3 (Brown et al., 2020) indicate that word embeddings—vector representations that encode the semantic meaning of words—can predict the perceived personality of public figures with considerable accuracy. This predictive ability persists even after statistically controlling for variables such as likeability and demographic information, suggesting that LLMs can extract specific information related to personality traits from their training corpora (Park et al., 2022).

These word embeddings not only possess predictive value but also exhibit an internal structure that reflects personality perception (Cao & Kosinski, 2024). Moreover, the models generated by these LLMs demonstrate high face validity, with descriptive personality adjectives appearing at the extremes of the dimensions predicted by the model (Atari et al., 2023; Wach et al., 2023). This finding reinforces the notion that the internal semantic structure of LLMs captures relevant dimensions of human personality, which can be applied in contexts such as marketing (Lee et al., 2024; Zhou et al., 2023; Demzky et al., 2020).

These findings open the possibility of using LLMs for message personalisation (Lee et al., 2024). Having captured information about an individual's personality or archetype, an LLM could tailor a message's style, tone, and content to maximise its persuasive impact (Matz et al., 2024). As quoted earlier, this ability has implications in advertising, marketing, political communication, and the development of personalised virtual assistants, among others (Peres et al., 2023).

The approach to these capabilities should be cautious. LLMs learn from vast textual datasets; thus, there is a risk that they reproduce and amplify biases present in training data, including biases related to gender, race, or cultural stereotypes (Bolukbasi et al., 2016; Campbell et al., 2023). Therefore, rigorous validation of the accuracy, reliability, and fairness of personality predictions generated by LLMs is crucial before their implementation in real-world applications (Lee & Kim, 2024). In addition, using LLMs for message personalisation raises important ethical issues related to manipulation, privacy, and transparency (Lee & Kim, 2024). One approach to bias mitigation is 'bias calibration prompts', which aim to rebalance model outputs by enforcing fairness constraints. Others include fine-tuning curated, debiased datasets or employing post-processing filters (Zhou et al., 2023).

When LLMs infer personality traits, moral values, or ideological leanings to tailor content, there is a risk that personalisation can cross into psychological manipulation. For example, hyper-personalised political messages or persuasive ads targeting emotional vulnerabilities may exploit cognitive biases, nudging consumers toward decisions that do not align with their true preferences or interests (H. Wang & Lu, 2025). Such practices blur the line between persuasion and coercion, undermining consumer autonomy. Practitioners and platform designers should implement transparency measures, such as informing users when content is personalised, disclosing the use of inferred psychological profiles, and allowing opt-out options, to foster trust and ensure that

personalisation respects individual agency and data ethics (Anthis et al., 2025; Yenduri et al., 2024).

### LLMs and market research

Traditionally, brand perception studies have relied on consumer surveys to collect data on consumer opinions and preferences. However, the irruption of LLMs has opened new avenues for market research, offering alternative methods with the potential to increase the efficiency and scalability of these studies. In this context, two recent papers have explored the application of LLMs to brand perception analysis: Li et al. (2024) and Brand, Israeli and Ngwe (2023).

Li et al. (2024) propose an automated method for brand perception analysis based on response generation from carefully designed prompts. Their methodology relies on the ability of LLMs to complete sentences and generate coherent text within a given context. For example, they use prompts such as 'Car Brand X is similar to Car Brand...' to assess perceived similarity between car brands. The responses generated by the LLM are further processed using text analysis techniques to obtain similarity scores and construct perception maps that visualise the relationships between brands in a multidimensional space.

Brand, Israeli and Ngwe (2023) evaluate the usefulness of GPT for simulating human responses in marketing research studies, focussing specifically on estimating the willingness to pay (WTP) for different attributes of a product. To this end, they employ GPT to simulate responses to conjoint analysis questions, a common technique in marketing research to extract consumer preferences and quantify the relative importance of different product attributes (Green & Srinivasan, 1978). In this context, GPT has a choice scenario where it must select between two product configurations with different attributes and prices, simulating a consumer's decision-making process. From these simulated choices, WTP is estimated for each product attribute.

Both methodologies depend on strategically using prompts to obtain relevant information from the LLM. However, they differ in the type of data collected and the objective of the analysis. Li et al. (2024) focus on assessing brand similarity and attribute scoring, using prompts that induce the LLM to complete sentences related to perceived similarity or to identify the brand that best matches a specific attribute. In contrast, Brand, Israeli and Ngwe (2023) focus on estimating WTP, using conjoint analysis-type questions to simulate a choice scenario between different product configurations to infer underlying consumer preferences.

Both studies compare the results obtained from LLM with empirical data collected through consumer surveys to validate the effectiveness of their respective methodologies. Li et al. (2024) introduce a novel 'triplet method' to compare LLM-generated and human-generated perception data sets by assessing the agreement between similarity relationships and attribute scores obtained by both methods. In turn, Brand, Israeli and Ngwe (2023) compare WTP estimates obtained from GPT-simulated responses with estimates derived from the responses of human participants in traditional conjoint analysis studies.

Their results suggest that LLMs can be a valuable tool for market research, providing relevant information on consumer perceptions. However, there are inherent limitations to these models. Li et al. (2024) caution that the nature and scope of the training data limit the capability of LLMs, which may introduce biases in the results. Similarly, Brand, Israeli and Ngwe (2023) note that while GPT can simulate responses at the aggregate level, it may have difficulty capturing the heterogeneity inherent in individual consumer preferences.

Park et al. (2022) take a different approach, introducing 'social simulations' to prototype social computing systems. Although not directly focussed on consumer psychology, the proposal by Park et al. (2022) allows them to take advantage of their ability to simulate human interactions in digital environments. Applied to a product forum, for example, an LLM could generate realistic conversations between consumers with diverse profiles, allowing the analysis of the formation of perceptions, the dissemination of information, and the emergence of

opinion leaders. This simulation would facilitate the identification of consumption patterns, the deep understanding of needs and desires, and the design and testing of marketing strategies, such as advertising campaigns or responses to reputational crises, before their actual implementation.

Our research question is as follows: **To what extent can A) models, specifically LLMs, replace or complement TMR (Market Research) methods in the evaluation of brand perception and consumer psychology?** Following Anthis et al. (2025), ‘complement’ here refers to the use of LLMs as auxiliary tools that support or enhance TMR methods, such as assisting in survey design, generating hypotheses, or simulating responses during exploratory phases, without displacing human respondents or analysts. By ‘replace’, we imply a full substitution of traditional methodologies with AI-generated data or analysis, treating LLMs as standalone research agents. **To answer this question, we set the following three hypotheses:**

**Hypothesis 1. LLMs have the potential to complement and even replace human involvement in market research.**

One of LLMs’ promises is that they respond similarly to humans. Thus, they could replace or complement TMR methods (Kshetri, 2023, 2024). Previous studies support the general idea that LLMs have great potential in market research, both as a complement and a possible substitute for specific tasks. For instance, Singh, Kumar and Mehra (2023) validated GPT-3 responses against Nielsen panel data with 74 % alignment in attribute prioritisation. Our work complements studies to determine the validity and reliability of LLMs in different market research contexts and to establish best practises for their use (Hermann & Puntoni, 2024; Kshetri et al., 2024).

Thus, Arora et al. (2024) demonstrate the potential of LLMs as active collaborators in qualitative research. Their ‘human-IA hybrid’ model suggests that LLMs can assist in question generation, topic identification, and preliminary data analysis, optimising the researcher’s role toward interpretation and insight generation tasks. This approach emphasises synergy rather than complete substitution.

Li et al. (2024) evidence the ability of LLMs to perform automated perceptual analysis with high concordance over traditional survey-based methods. Their work leans toward task-specific substitution, arguing for the efficiency and accuracy of LLMs in generating perceptual maps from brand and attribute similarity data.

Sarstedt et al. (2024) take a more cautious stance on substituting human participants for *silicon samples* generated by LLMs. While acknowledging the potential of LLMs to generate synthetic data, a review of the existing literature reveals mixed results compared to human data, especially in contexts that demand a deep understanding of cognition and emotions. They recommend an exploratory use in the preliminary phases of research.

Our first objective is to determine the extent to which traditional surveys can be made possible through GAI. Our work enriches the existing literature by comparing different LLM architectures, supporting the second hypothesis.

**Hypothesis 2. LLMs can capture and reflect aspects of consumer psychology, including perceptions of brand personality, in a manner comparable to TMR methods.**

This capability derives from their ability to process natural language, learn complex patterns from vast data sets, and infer psychological constructs from linguistic expressions, thus allowing them to emulate human perception of brands. One of these areas is the ability to identify brand personality traits. Using the self-verification theory, Park and Ahn (2024) show that GAI—in this case, ChatGPT—can identify luxury brand personality traits that align with consumers’ self-concept. While traditional models exhibit greater explanatory power, AI-generated content demonstrates significant potential to reflect brand personality perceptions, encouraging brand identification.

Brand, Israeli and Ngwe (2023) demonstrate that LLMs (GPT-3.5

Turbo) can generate responses to market research surveys that reflect what consumers would give. Their study on WTP shows that estimates obtained from GPT responses are realistic and comparable to those from human studies, indicating the ability of LLMs to capture consumer preferences.

AI can also aid in direct consumer engagement by creating Personalised Persuasive Messages tailored to psychological profiles (personality traits, ideologies, moral underpinnings), which are more effective than non-personalised ones. LLMs can infer and use psychological constructs to influence consumer perception and behaviour (Campbell et al., 2023; Lee et al., 2024; Matz et al., 2024).

Generating Data from *silicon samples* with Limitations: Sarstedt et al. (2024) examine the use of LLMs to generate *silicon samples*. While acknowledging limitations in fully mimicking human behaviour and incredibly complex cognitive processes, they suggest that LLMs may be valuable in the early stages of research or for specific tasks, suggesting a potential for understanding aspects of consumer psychology with appropriate development.

Li et al. (2024) demonstrate that LLMs can generate human-like responses in brand perception surveys, enabling the creation of accurate perceptual maps and the exploration of consumer heterogeneity. They suggest that LLMs trained on vast data sets can overcome human limitations in understanding and utilising psychological constructs, tailoring messages to individual profiles with greater accuracy (Li et al., 2024).

LLMs detect psychological cues primarily through the linguistic analysis of adjectives, tone, syntactic structure, and contextual semantics within user-generated content. When exposed to text rich in evaluative language (e.g., ‘refreshing’, ‘pretentious’, ‘authentic’), models such as GPT-4 can infer dimensions of personality using pre-trained associations between linguistic features and psychological traits. Atari et al. (2023) and Cao and Kosinski (2024) demonstrate that LLMs can predict perceived personality traits of brands and individuals with surprising accuracy, even outperforming humans in some tasks of attribute matching. These capabilities allow LLMs to generate marketing messages tailored to personality archetypes, increasing persuasive effectiveness and engagement.

**Hypothesis 3. The value of the results generated by LLMs in market research decreases as the level of concreteness of the questions asked increases because of the difficulty of LLMs to process and generate accurate and relevant information about specific products or brands compared to general or abstract information.**

This hypothesis—a contribution to the literature—is based on inferences drawn from the existing literature on the capabilities and limitations of LLMs in market research-related contexts. Sarstedt et al. (2024) observe that GPT emphasises high-level abstraction and goal-oriented features to the detriment of low-level and means-oriented features, suggesting a remarkable aptitude of LLMs to handle general and abstract information, which would imply lower effectiveness in addressing product—or brand-specific details.

In parallel, Li et al. (2024) find that these models’ contribution is limited in the initial stages that require specific market knowledge, such as defining the breadth of a category or selecting representative brands. Finally, Brand, Israeli and Ngwe (2023) warn about the possibility of LLMs generating erroneous conclusions when evaluating new or particular product characteristics. They illustrate with examples the difficulty of LLMs handling product-specific information reliably, which could compromise the usefulness that Hypothesis 5 proposes, that is, the diverse construction of GAI models can cause substantial variations in market research results. Researchers should be aware of such heterogeneity and carefully consider the choice of LLM when designing and executing market research studies. Comparing results between different LLMs and testing the robustness of the findings are essential steps to ensure the validity and generalisability of the conclusions (Costa & Di Pillo, 2024; Peng, 2023).



Sarstedt et al. (2024) investigate using LLMs to generate *in silicon samples* in market research and highlight that results vary considerably across domains and models used. The authors attribute this variability to differences in the operating principles of LLMs, including ways to customise and parameterise them. They also point to researchers' use of different versions of LLMs as another source of variability in the results.

Brand, Israeli, and Ngwe (2023) test the robustness of the results as a function of the LLM chosen. They replicated their study using several OpenAI, Meta, and Anthropic models. They found notable differences in responses, such as WTP ratings for the MacBook brand and screen size preference. Based on these findings, the authors argue that researchers should be aware of the sensitivity of the results to the choice of LLM. Li et al. (2024) suggest that the choice of language model may influence the results. Although the authors focus on two specific models (GPTNeo 2.7B and ChatGPT based on GPT4), they suggest their method can be adapted to any generative model. Szczesniewski et al. (2024) obtain similar results in a different setting, reinforcing the relevance of the analysis of the hypotheses.

LLMs show limitations in handling brands when analysing emergent or niche markets with limited digital presence. For example, Brand, Israeli and Ngwe (2023) use GPT-3.5 to estimate WTP for new product attributes in a tech accessories market. While the model accurately predicts preferences for widely known features (e.g., battery life or screen size), it does not yield consistent responses for a lesser-known brand. The model overestimated brand recognition and assigned implausibly high WTP values, likely because of a lack of training data for the brand in question.

## Data collection, study design and execution

The research aims to determine to what extent TMR can be replicated by using GAI. For this purpose, we have designed and executed a real market research survey conducted by the professional service provider MásMétrica, focussing on the beer market industry in Spain. This survey provides a deep understanding of beer consumption habits in Spain, covering all the stages of the conversion funnel, from Awareness and Consideration of the Brand to Purchase and Consumption Experience. It also includes the evaluation of specific attributes most valued by the customer in relation to two specific well-known beer brands: Heineken and Corona. Then, we compare these results with the responses generated for the same questions by four of the most popular LLM models currently available in the market: ChatGPT from OpenAI, Gemini from Google, Claude from Anthropic, and LLaMa from Meta.

We have chosen the beer market for this research because it not only allows us to investigate customer behaviour in a well-defined, data-rich context but also serves as a compelling case to compare TMR methods with AI-driven approaches.

The beer market offers an ideal testing ground for comparing TMR with AI because of its complexity, data availability, practical relevance, and the opportunity to effectively address specific challenges that AI could solve (Nave et al., 2022). The beer market is mature, with a high concentration of players and intense competition, which generates a large amount of historical and current data helpful for training and evaluating AI models (Malone & Lusk, 2018; Ricciuto et al., 2006). Despite its maturity, the sector is constantly evolving because of factors such as the growing popularity of craft beers, consumer trends toward premium and non-alcoholic products, and innovation in flavours, formats, and production processes (Thomé et al., 2017). These conditions create a scenario for comparing traditional research and AI's adaptive and predictive capabilities.

### Traditional market research survey (TMRS): study design, setting and distribution

A cross-sectional survey was conducted by MásMétrica, a professional market research services provider. The objective is to provide a

contemporary overview of consumer knowledge, opinions, perceptions, and preferences about beer consumption in Spain. The universe and population under study are people aged 18 to 75 years, residing in Spain at the time of the survey.

In the present study, a sample of 1003 individuals aged 18 to 75, belonging to a panel associated with MásMétrica, was selected intentionally and proportionally to the quotas of the Spanish population (according to the provisional INE Population Census as of 1 January 2024), based on sex, age group, and Autonomous Community (CCAA).

The questionnaire was self-administered via email through a link that was enabled from 24 to 26 July 2024. The final sample size (n) is 1003, with a maximum sampling error of  $\pm 3.1\%$  at a 95 % confidence level, assuming  $P = Q = 50\%$  (maximum variability). Sample size and level of precision are standard for national surveys conducted by national statistical institutions.

### Survey design

The survey comprised 24 items, 9 focusing on demographics and 15 on consumer knowledge, opinions, perceptions, and preferences regarding beer consumption. The questions were based on previous studies' validated questionnaires on beer consumption (Malone & Lusk, 2018; Vrontis, 1998; Wang et al., 2023). Descriptive statistics of the demographic characteristics of the sample in the TMRS are provided in Tables A1-A9 in Annexe A. To ensure a concise yet comprehensive analysis, we strategically selected six key questions that cover all stages of the conversion funnel (Fig. 1) (Colicev et al., 2019). These questions are crucial for assessing customer behaviour from awareness to experience and were chosen for their established relevance and significant contribution to understanding the dynamics of beer consumption, as established in previous research (Vrontis, 1998; Malone & Lusk, 2018), without compromising the depth and validity of the insights.

### LLM market research: study design and execution

In parallel, the same survey was conducted using four of the most prominent LLMs models available on the market: ChatGPT by OpenAI, Gemini by Google, Claude by Anthropic, and LLaMa by Meta (Table 1).

To conduct our LLM market research, we developed a methodological roadmap to ensure the robustness of the research process (Fig. 2). We asked the LLM models to perform as market researchers and prompted them with the questions from the original market research questionnaire. This step was crucial to directly compare LLMs' responses with those of the TMRS. We then requested the models to identify the sources used to generate their responses, to ensure transparency and traceability of the generated data. For models without internet access, we required them to rely on their pre-trained databases, simulating a real-world scenario. Additionally, for multiple-choice questions, we ensured the model could accurately recognise and respond to these question formats by instructing it to correct its response when necessary. This methodological roadmap ensures that LLM-generated responses are rigorously produced and are comparable with those from TMRS.

In this research, we employ widely recognised LLM models currently available in the market (Table 1). However, to preserve scientific rigour and ensure neutrality, specific model names will be anonymised throughout the article, as the objective is not to compare commercial performance but to provide unbiased insights.

## Empirical analysis and results

This section introduces the empirical analysis and results. First, it presents a description of the statistical methods used to compare AI-based survey responses with those from our TMRS, specifically within the beer market in Spain. Second, the results are explained through the comparison of frequency distributions, the testing of hypotheses regarding their equality, and measuring the degree of similarity across various beer consumption questions to assess how closely the responses

**Table 1**  
LLMs used to simulate Market Research: Case Study on Beers in Spain.

Model	Version	Type	Base Architecture	Training	Main Focus	Multimodality	Key Features
ChatGPT (OpenAI)	ChatGPT4o	LLM	Transformer	Pre-trained on general and specific text data	Text generation, conversational assistance, business, and personal applications.	Yes (GPT-4 V processes text and images).	Advanced conversational capabilities, widely accessible. Pioneer in providing public access to LLMs
Gemini (Google)	Gemini Advanced	Multimodal Model (LLM)	Transformer	Trained on textual, image, video and audio data	Advanced multimodality, text generation, image analysis, and understanding of audio-visual content.	Yes (text, images, videos, and audio).	Pioneering multimodality, deeply integrated with Google services
Claude (Anthropic)	1.5Pro Claude 3.5	LLM	Transformer	Focused on ethical alignment and robustness	Text generation, business task assistance, designed for ethical and safe usage.	Yes (Claude 2 with limited capabilities).	Prioritises ethics and safety, alignment-focused training
Llama (Meta)	Llama 3.2	LLM	Transformer	Trained on diverse, high-quality text datasets	Academic research and advanced NLP applications.	No	Optimised for hardware efficiency, designed for research

Compiled in-house, based on the following sources: (Anil et al., 2023; Bai et al., 2022; OpenAI, 2023; Touvron et al., 2023).

**Table 2**  
Comparison of each LLM with the TMRS: Chi-Square Test Results and Similarity Metrics.

Compared Model	Chi-square	p-value	Degrees of Freedom	Jensen-Shannon Distance	MAD
LLM_1	294.28	6.80E-59	8	0.2845	69.89
LLM_2	374.65	4.93E-76	8	0.3304	82.11
LLM_3	256.88	6.00E-51	8	0.2625	64.78
LLM_4	78.13	1.16E-13	8	0.1415	32.33

**Table 3**  
Chi-square test of the joint distribution of TMRS and the four LLMs for each Corona brand attribute.

Brand Attributes: Corona	Chi-square	p-value	Degrees of Freedom
Price	474.63	7.46E-91	16
Flavour	931.53	5.01E-188	16
Advertising	672.67	8.40E-133	16

**Table 4**  
Comparison of each LLM with the TMRS: Chi-Square Test Results and Similarity Metrics for three key attributes of the Corona Brand.

Attribute: Price					
Compared Model	Chi-square	p-value	Degrees of Freedom	Jensen-Shannon Distance	MAD
LLM_1	73.67	3.82E-15	4	0.1449	52.40
LLM_2	76.79	8.31E-16	4	0.1466	68.40
LLM_3	53.09	8.17E-11	4	0.1227	35.20
LLM_4	233.63	2.19E-49	4	0.2594	120.80
Attribute: Flavour					
Compared Model	Chi-square	p-value	Degrees of Freedom	Jensen-Shannon Distance	MAD
LLM_1	429.56	1.14E-91	4	0.3541	171.20
LLM_2	81.66	7.74E-17	4	0.1510	68.00
LLM_3	37.20	1.64E-07	4	0.1017	46.00
LLM_4	357.58	4.06E-76	4	0.3228	151.20
Attribute: Advertising					
Compared Model	Chi-square	p-value	Degrees of Freedom	Jensen-Shannon Distance	MAD
LLM_1	148.92	3.47E-31	4	0.210	74.40
LLM_2	146.06	1.42E-30	4	0.206	90.80
LLM_3	84.88	1.61E-17	4	0.156	45.20
LLM_4	305.03	8.92E-65	4	0.306	111.20

**Table 5**

Chi-square test of the joint distribution of TMRS and the four LLMs for each Heineken brand attribute.

Brand Attributes: Heineken	Chi-square	p-value	Degrees of Freedom
<b>Price</b>	829.26	3.60E-166	16
<b>Flavour</b>	489.35	5.86E-94	16
<b>Advertising</b>	775.37	1.13E-154	16

**Table 6**

Comparison of each LLM with the TMRS: Chi-Square Test Results and Similarity Metrics for three key attributes of the Heineken Brand.

Attribute: Price					
Compared Model	Chi-square	p-value	Degrees of Freedom	Jensen-Shannon Distance	MAD
LLM_1	61.80	1.21E-12	4	0.1295	38.40
LLM_2	267.92	8.97E-57	4	0.2687	132.80
LLM_3	79.96	1.78E-16	4	0.1464	64.40
LLM_4	230.62	9.73E-49	4	0.2474	131.60
Attribute: Flavour					
Compared Model	Chi-square	p-value	Degrees of Freedom	Jensen-Shannon Distance	MAD
LLM_1	38.97	7.07E-08	4	0.1020	38.80
LLM_2	163.99	2.04E-34	4	0.2087	105.60
LLM_3	10.88	2.79E-02	4	0.0532	18.40
LLM_4	111.83	2.96E-23	4	0.1714	89.20
Attribute: Advertising					
Compared Model	Chi-square	p-value	Degrees of Freedom	Jensen-Shannon Distance	MAD
LLM_1	204.15	4.81E-43	4	0.2373	91.60
LLM_2	353.17	3.62E-75	4	0.3102	158.00
LLM_3	54.58	3.97E-11	4	0.1204	38.40
LLM_4	150.32	1.74E-31	4	0.2000	90.80

**Table 7**

Comparison of each LLM with the TMRS: First-order Rao-Scott corrected chi-square test results and Similarity Metrics.

Compared Model	First-order Rao-Scott Corrected Chi-square	p-value	Degrees of Freedom	JSD	MAD
LLM_1	365.19	0.00	10	0.2400	126.27
LLM_2	523.84	0.00	10	0.3283	175.18
LLM_3	395.14	0.00	10	0.2104	102.82
LLM_4	198.64	0.00	10	0.2141	167.73

generated by each LLM method match those obtained through the TMRS. Finally, a summary of the findings is presented.

#### Statistical methods

This research focuses on six questions from the market research survey, which include both single-response and multiple-response

**Table 8**

Comparison of each LLM with the TMRS: Chi-Square Test Results and Similarity Metrics.

Compared Model	Chi-square	p-value	Degrees of Freedom	Jensen-Shannon Distance	MAD
LLM_1	177.94	2.07E-37	4	0.2131	112.60
LLM_2	254.21	8.05E-54	4	0.2554	135.40
LLM_3	66.19	1.44E-13	4	0.1289	56.20
LLM_4	86.89	6.02E-18	4	0.1477	75.40

**Table 9**

Comparison of each LLM with the TMRS: First-order Rao-Scott corrected chi-square test results and Similarity Metrics.

Compared Model	first-order Rao-Scott Corrected Chi-square	p-value	Degrees of Freedom	JSD	MAD
LLM_1	562.62	0.0000E+00	3	0.3144	332.00
LLM_2	50.57	6.0489E-11	3	0.0904	77.50
LLM_3	78.14	1.1102E-16	3	0.1088	56.00
LLM_4	296.99	0.0000E+00	3	0.2167	178.75

formats corresponding to different stages of the conversion funnel that customers—in this case, beer customers—experience.

The selected statistical tests and measures are appropriate for analysing survey questions with these specifications, enabling a methodological comparison of categorical distributions from complementary perspectives: first, hypothesis testing for the simultaneous comparison of the five distributions, including the TMRS and the responses from multiple LLMs and second, two additional metrics for one-to-one comparisons between the TMR responses and each individual LLM-generated response.

This methodological approach is grounded in the theoretical assumption that LLMs can act as proxies for human judgment in customer behaviour analysis and market research. Accordingly, the statistical techniques selected aim to validate this premise by comparing the structure and consistency of the LLMs-generated responses with those from TMRS.

#### Hypothesis testing of distribution equality across multiple populations

The **chi-square** distribution test of homogeneity of proportions is suitable for working with categorical single-response questions. The aim is to test the null hypothesis that no difference exists in the proportion of responses in a set of mutually exclusive categories between two or more populations (Garson & Moser, 1995).

$$\chi^2_{(F-1)(C-1)} = \sum_{i=1}^F \sum_{j=1}^C \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (1)$$

Where  $O_{ij}$  are the observed frequencies, and  $E_{ij}$  are the expected frequencies under the null hypothesis.  $F$  represents the number of populations (market research methods' responses), and  $C$  denotes the number of mutually exclusive categories for each question. We first use this test to compare the six distributions simultaneously and later to perform four one-to-one tests comparing each LLM-generated response with the TMR results for each question. In the case of categorical multiple-response questions, the same approaches are used, but the first-order Rao-Scott corrected chi-square is applied instead. Hence, the correction factor  $\hat{c}$  minimises the impact of dependencies that may arise from multiple responses, where  $\hat{c}$  is the ratio of expected variance under independence to the observed variance (Bilder & Loughin, 2001; Heo,

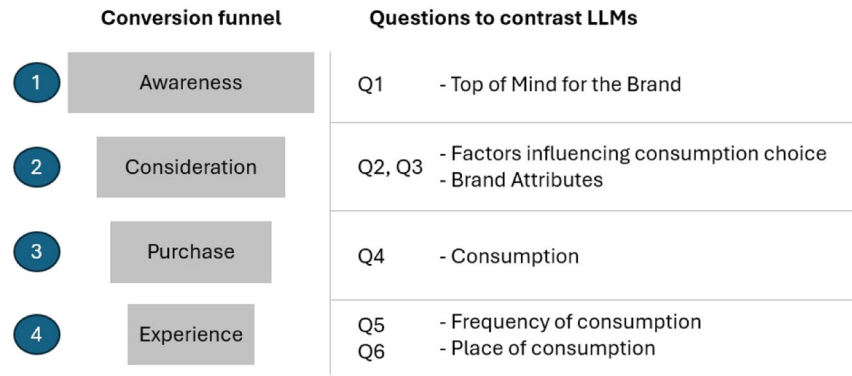


Fig. 1. Conversion Funnel Stages and their corresponding questions to contrast LLMs.

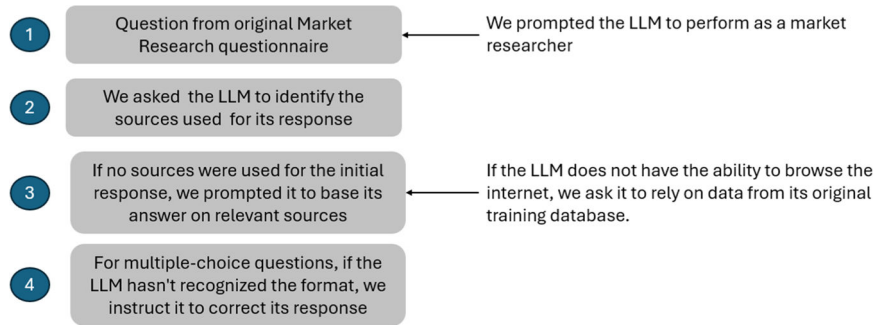


Fig. 2. Methodological Roadmap for LLM Market Research.

2006; Lavassani et al., 2009).

$$\chi_{RS}^2 = \frac{\chi^2}{c} \quad (2)$$

#### Measuring distribution similarity between two populations

We use the Jensen-Shannon Distance (JSD) statistic to measure the similarity between the distributions of two populations—in this case, between each LLM and the TMRS. The goal is to determine to what extent any of the LLM-generated responses can replicate the distribution of those provided for the TMRS, as well as to facilitate the ranking of the AI models' performance. This metric (JSD) is a symmetrised, smoothed version of the Kullback-Leibler divergence, which measures the relative difference between two probability distributions (Pinto et al., 2018; Sadrani et al., 2024).

Specifically, given two distributions  $P$  and  $Q$ , the divergence is defined as follows:

$$JSDiv(P, Q) = \frac{1}{2}D_{KL}(P \parallel M) + \frac{1}{2}D_{KL}(Q \parallel M) \quad (3)$$

where  $M = \frac{1}{2}(P + Q)$  is the mean distribution.

The JSD is obtained by taking the root of the Jensen-Shannon Divergence as follows:

$$JSD(P, Q) = \sqrt{JSDiv(P, Q)} \quad (4)$$

This metric is highly robust for categorical distributions, including both single-response and multiple-response questions. Its value ranges from 0 to 1, where 0 represents identical distributions, and values close to 1 indicate significant differences between the distributions (Briët & Harremoës, 2009; Connor et al., 2013). Values below 0.1 indicate very high similarity, while those below 0.5 indicate sufficient similarity. Values above 0.5 indicate high dissimilarity (Blair et al., 2012; Spinner et al., 2021).

#### Measuring the average pointwise discrepancy between two populations

The Mean Absolute Difference (MAD) is used to measure the similarity between two vectors. This statistic quantifies the average absolute error between two populations—in this case, the difference between the responses provided by each of the AI models and the TMRS. This metric is widely adopted for evaluating the accuracy of prediction models. If  $A = (a_1, a_2, \dots, a_n)$  and  $B = (b_1, b_2, \dots, b_n)$  are vectors of size  $n$  (Song et al., 2003; Ye et al., 2024), the MAD is defined as follows:

$$MAD = \frac{1}{n} \sum_{i=1}^n |a_i - b_i| \quad (5)$$

#### Empirical analysis and results

In this section, we provide a descriptive statistical analysis comparing the relative frequency distribution between the TMRS and the responses generated by LLMs across different beer consumption questions (Annexe B). Additionally, we apply the statistical methods described in Section 4.1 to test equality hypotheses and measure the similarity between the responses obtained from the TMR and those generated by each LLM across all the different questions. Finally, a summary of the results and findings is presented.

##### Awareness stage in the conversion funnel: survey questions focused on user awareness and spontaneous brand recall

At the awareness stage of the conversion funnel, the research focuses on one question related to the first brand that comes to the customer's mind when asked about beer brands.

##### Question 1: If we talk about beer brands, which beer brands do you know or come to mind?

The results of the TMRS show not only a high level of awareness of several brands but also market fragmentation, with high variability in the responses. This question has been processed as a single response using the first brand mentioned by the 1003 respondents. Mahou is the



most frequently mentioned brand by beer consumers, with 24.53 % of the mentions, followed by Estrella Galicia with 13.86 % and Heineken with 10.57 %. Other brands mentioned include Cruzcampo (8.47 %), Estrella Damm (7.68 %), San Miguel (5.48 %), Alhambra (4.69 %), and Amstel (4.19 %). Notably, 20.54 % of the customers mentioned other brands. Annexe B: [Table B1](#) shows the distribution of the relative frequency of responses regarding top-of-mind beer brands among customers, including a comparison with LLM-generated responses.

The chi-square test shows significant differences among the five response distributions (TMRS and four LLMs) with a value of 1541.06, degrees of freedom (df) = 32, and p-value = 3.59e-304, leading to the rejection of the null hypothesis of distribution equality.

Complementarily, four chi-square tests were performed to compare the response distributions of each LLM with the TMRS ([Table 2](#)). All the p-values were smaller than 0.01, leading to the rejection of all four null hypotheses. This means that none of the LLMs can fully replicate the response distribution of the TMRS regarding beer awareness and spontaneous brand recall, indicating significant differences.

Nevertheless, we observe differences in the performance of the LLMs. The model with the smallest chi-square value is LLM\_4 (chi-square = 78.13), indicating its responses are the most aligned with those of the TMRS.

This finding is also supported by the JSD and the MAD metrics, indicating high similarity for the LLM\_4 (JSD = 0.1415 and MAD = 32.33), lower than 0.2 and 50, respectively, suggesting that it can serve as a reliable proxy for estimating the beer awareness and spontaneous brand recall in Spain. The other LLMs show lower similarity to TMRS, with JSD higher than 0.2, but still below the cut-off of 0.5, and MAD values greater than 50.

Therefore, H1 is accepted, particularly for LLM\_4, as its responses, although they cannot fully replicate those provided by the TMRS, can serve as a proxy and complement them. As this question corresponds to the awareness stage of the conversion funnel, the variability observed in brand recall responses may affect strategic decisions made by decision-makers regarding brand positioning, market share estimations, and top-of-mind visibility in the Spanish beer market. The closer alignment of LLM\_4 with TMRS results suggests its potential as a complementary tool for identifying leading brands in consumers' minds and detecting fragmentation in spontaneous recall patterns. This stronger alignment may be related to the specific characteristics of the model that showed the closest results to TMRS, as reported in [Table 2](#). These findings may help inform the design and selection of models better suited for the awareness stage tasks in customer behaviour research.

#### *Consideration stage in the conversion funnel: survey questions focused on factors influencing consumption decisions and perceived value of specific products and brand attributes*

At the consideration stage of the conversion funnel, the research focuses on comparing specific attributes of two well-known beer brands, Heineken and Corona. Interviewees are asked to evaluate attributes such as price, flavour, and advertising in questions 2 and 3. These attributes were selected based on their examination in empirical studies and their identification as key drivers of brand evaluation, consumption, and purchase intention ([Chen et al., 2005](#); [Lerro & Nazzaro, 2020](#); [Meyerding et al., 2019](#); [Noel et al. 2018](#); [Orth, 2006](#)). The suitability of these attributes was also empirically supported by the results obtained in our TMRS as follows:

**Question 2: Evaluate the following attributes of the Corona brand: (1) price; (2) flavour; (3) advertising.**

**Question 3: Evaluate the following attributes of the Heineken brand: (1) price; (2) flavour; (3) advertising.**

The results of the TMRS show that the most significant factor influencing beer brand choice is flavour, with 82.3 % of the respondents citing it as their primary consideration. Price is the second most important factor, influencing 46.4 % of the consumers, followed by brand importance (40.2 %). Other factors include the availability at the

place of consumption or purchase (35.7 %), recommendations from friends or family (13.0 %), and advertising, which influences 6.2 % of the respondents.

Heineken is a highly valued brand in terms of flavour and, particularly, in advertising, with a vast majority of the respondents rating these attributes as very good or good (77.18 % and 82.47 %, respectively) and only a few of them as very bad or bad (15.46 % and 7.88 %, respectively). Annexe B: [Table B3](#) shows the distribution of the relative frequency of responses regarding the customers' perceived value of these attributes for the Heineken brand. In comparison, Corona gets slightly lower ratings, with most of the respondents rating the same attributes (flavour and advertising) as very good or good (74.42 % and 65.60 %, respectively) and only a few of them as very bad or bad (15.21 % and 16.76 %, respectively). Annexe B: [Table B2](#) shows the distribution of the relative frequency of responses regarding the customers' perceived value of these attributes for the Corona brand.

Customers tend to be more critical when evaluating attributes such as price. Heineken receives a 67.63 % rating for good and very good, but simultaneously, 23.24 % of the customers assess it as bad or very bad. The same effect happens with Corona, as 56.56 % of the customers rate its price as very good or good, but 30.76 % assess it as very bad or bad.

#### **Evaluating the attributes of the Corona brand**

Three chi-square tests of the joint distribution of the TMRS and the four LLMs for each Corona brand attribute ([Table 3](#)) show significant differences across their distributions, with p-values below 0.01. This leads to the rejection of the null hypothesis of distribution equality for the three Corona brand attributes: price, flavour, and advertising.

Complementarily, four chi-square tests were performed for each of the three Corona attributes (price, flavour, and advertising) to compare the response distributions of the LLMs with the TMRS ([Table 4](#)). All the p-values were smaller than 0.01, leading to the rejection of the 12 null hypotheses. This means that none of the LLMs can fully replicate the response distributions of the TMRS regarding the evaluation of the key attributes of the Corona brand, indicating significant differences.

Nevertheless, we observe differences in the performance of the LLMs, with a clear pattern of superior performance across the three key attributes. Among them, LLM\_3 stands out as the best performer, achieving the lowest chi-square values across all attributes (53.09, 37.20, and 84.88 for price, flavour, and advertising, respectively), indicating that its responses are the most closely aligned with those of the TMRS. LLM\_2 follows as the second-best model for the attributes of flavour and advertising, with chi-square values of 81.66 and 146, respectively, while LLM\_1 is the second-best for the price attribute, with chi-square values of 73.67.

This finding is also supported by the JSD and MAD metrics, indicating high similarity for LLM\_3 with JSD values of 0.1227, 0.1017, and 0.156, all lower than 0.2, and MAD values of 35.20, 46.00, and 45.20, all lower than 50, for price, flavour, and advertising, respectively. These metrics suggest that LLM\_3 responses can serve as a reliable proxy for estimating the customers' perceptions about these three key attributes.

LLM\_1 could serve as a less reliable proxy for the price attribute (JSD = 0.1449 and MAD = 52.40), and LLM\_2 could serve the same role for the attributes of flavour and advertising (JSD values of 0.1510 and 0.206, and MAD values of 68.00 and 90.80).

#### **Evaluating the attributes of the Heineken brand**

Three chi-square tests of the joint distribution of TMRS and the four LLMs for each Corona brand attribute ([Table 5](#)) show significant differences across their distributions, with p-values below 0.01. This leads to the rejection of the null hypothesis of distribution equality for the three Heineken brand attributes: price, flavour, and advertising.

Complementarily, four chi-square tests were performed for each of the three Heineken attributes (price, flavour, and advertising) to compare the response distributions of the LLMs with the TMRS ([Table 6](#)). All the p-values were smaller than 0.01, leading to the rejection of the 12 null hypotheses. This means that none of the LLMs can fully replicate the response distributions of the TMRS regarding the

evaluation of the key attributes of the Heineken brand, indicating significant differences.

Nevertheless, we observe differences in the performance of the LLMs, with performance varying by beer attribute. Among them, LLM\_3 stands out as the best performer for two out of three attributes. It achieves the lowest chi-square values across the flavour and advertising attributes (10.88 and 54.58, respectively), indicating that its responses are the most closely aligned with those of the TMRS for these two attributes. LLM\_1 stands out as the best performer for the price attribute, with a chi-square value of 61.80, followed by LLM\_3 with 79.96. The second-best performing models for the attributes flavour and advertising are LLM\_1 and LLM\_4, with chi-square values of 38.97 and 150.32, respectively.

These findings are also supported by the JSD and the MAD metrics, indicating high similarity for LLM\_3 in the beer attributes of flavour and advertising (JSD = 0.0532 and 0.1204, lower than 0.2, and MSD = 18.40 and 38.40, both lower than 100), while LLM\_1 for the price attribute has JSD = 0.1295 and MAD = 38.40. LLM\_1—the second-best performing model in terms of similarity for flavour—has JSD = 0.1020 and MAD = 38.80; for advertising, it is LLM\_4, with metrics closer to the cut-off and with higher JSD = 0.20 and MAD = 90.80, but still below the acceptable thresholds of 0.5 and 100, respectively.

#### As a summary of the evaluation of attributes of the Corona and Heineken brands:

H1 is accepted, particularly for LLM\_3, as its responses for all three product attributes of the Corona brand and the attributes of flavour and advertising in the Heineken brand, although they cannot fully replicate those provided by the TMRS, can serve as a proxy and complement to them. For the price attribute of Heineken, LLM\_1 performed better, indicating that not all attributes are as well captured by a single model.

However, there is enough statistical evidence to accept H3, as LLMs can capture and reflect aspects of customer psychology—in this case, the tendency of customers of both Corona and Heineken to be more critical when evaluating attributes such as price—and the recognition of highly personal product preferences, such as flavour, as well as attitudes towards advertising (like or dislike).

These results also highlight the importance of combining several models to calibrate and reinforce the validity of some results when using LLMs to replicate TMRS.

#### *Purchase stage in the conversion funnel: survey questions focussed on users' brand consumption behaviour*

At the purchase stage of the conversion funnel, the research focuses on one question related to the users' habitual consumption behaviour.

**Question 4: Which beer brands do you usually consume? Select all that apply.**

(1) Mahou; (2) Estrella Galicia; (3) Heineken; (4) Cruzcampo; (5) Estrella Damm; (6) San Miguel; (7) Alhambra; (8) Amstel; (9) Águila; (10) Corona; (11) Others.

To analyse this question, consumers who stated that they did not drink beer were excluded, adjusting the sample size to 921 respondents. The customers are offered ten different brands of beer, along with the option to choose an additional 'others' category. This is a multiple-response question; thus, the sum of percentages exceeds 100 %.

The results of the TMRS show the presence of two highly popular brands, alongside significant market fragmentation, as reflected in the variability of responses. The most preferred brands are Estrella Galicia (49.84 %) and Mahou (42.35 %), followed by Heineken (30.62 %). The remaining brands are selected by <30 % of the respondents: Alhambra (27.14 %), San Miguel (25.41 %), Cruzcampo (23.45 %), and Estrella Damm (22.69 %). Finally, the less popular brands include Amstel (15.96 %), Águila (15.20 %), and Corona (15.20 %). Notably, 41.04 % of the respondents mentioned other brands, highlighting the diversity of preferences within the beer market. Annexe B: Table B4 shows the distribution of the relative frequency of responses regarding customers' brand consumption intention, including a comparison with LLM-

generated responses.

The first-order Rao-Scott corrected chi-square test shows significant differences between the five response distributions (TMRS and four LLMs) with a value of 1468.35, degrees of freedom (df) = 40, and p-value 0.0, leading to the rejection of the null hypothesis of distribution equality.

Complementarily, four first-order Rao-Scott corrected chi-square tests were performed to compare the response distributions of each LLM with the TMRS (Table 7). All the p-values were smaller than 0.01, leading to the rejection of all four null hypotheses. This means that none of the LLMs can fully replicate the response distribution of the TMRS regarding users' habitual beer brand consumption behaviour, indicating significant differences.

In this case, the LLMs show less variation compared to other survey questions. LLM\_2 has the smallest first-order Rao-Scott corrected chi-square with 198.64. However, LLM\_3 outperforms with the lowest JSD and the MAD metrics, with values of 0.2104 and 102.82, respectively. Although the JSD is higher than 0.2, it remains below the 0.5 threshold, indicating moderate alignment with the traditional survey responses. The MAD also exceeds 100.

Therefore, H1 is partially accepted, particularly for LLM\_3, as its responses, although they cannot fully replicate those provided by the TMRS, demonstrate a low-moderate level of reliability and can cautiously serve as a proxy and a complementary source of insights. Additionally, H3 is accepted, as the precision of the LLMs in market research decreases when dealing with highly specific questions about future purchase intentions, brand or product positioning, and market share in relation to the industry landscape. This information is more difficult to find on the internet or in public sources used to train these models, and complexity increases with multi-response questions.

#### *Experience stage in the conversion funnel: survey questions on user experience and general consumption habits*

At the experience stage of the conversion funnel, the research focuses on two questions related to general beer consumption habits, specifically the frequency and the places where it occurs.

**Question 5: How frequently do you consume beer? The possible response options are as follows: (1) Daily; (2) At least once a week; (3) At least once a month; (4) Less frequently; (5) Never.**

The results of the TMRS show that most respondents (48.90 %) consume beer at least once a week, 24.90 % daily, 10.50 % monthly, 7.60 % less frequently than monthly, and 8.20 % never. This question is a single response and has 1003 respondents. Annexe B: Table B5 shows the distribution of the relative frequency of responses regarding beer consumption frequency, including a comparison with the LLM-generated responses.

The chi-square test shows significant differences among the five (the TMRS and four LLMs) with a value of 377.68, degrees of freedom (df) = 16, and p-value 1.710e-70, leading to the rejection of the null hypothesis of distribution equality.

Complementarily, four chi-square tests were performed to compare the response distributions of each AI-based model with the TMRS (Table 8). All the p-values were smaller than 0.01, leading to the rejection of all four null hypotheses. That is, none of the LLMs can fully replicate the response distribution of the TMRS regarding the frequency of beer consumption, indicating significant differences.

Nevertheless, we observe differences in the performance of the LLMs. The models with the smallest chi-square values are LLM\_3 (chi-square = 66.19) and LLM\_4 (chi-square = 86.89), indicating that their responses are more aligned with those of the TMRS compared to the other LLMs.

This finding is also supported by the JSD and the MAD metrics, indicating high similarity for the LLM\_3 (JSD = 0.1289 and MAD = 56.20) and LLM\_4 (JSD = 0.1477 and MAD = 75.40). Both models have JSD values below 0.2 and MAD values below 100, suggesting they can serve as reliable proxies for estimating the frequency of beer consumption in Spain. In contrast, the other LLMs show lower similarity, with

JSD values greater than 0.2 but still below the cut-off of 0.5, indicating moderate alignment with the traditional survey responses and much higher MAD values (greater than 100).

Therefore, H1 is accepted, particularly for LLMs 3 and 4, as their responses, although they cannot fully replicate those provided by the TMRS, can serve as a proxy and complement them.

**Question 6: Where do you usually consume beer? Select all that apply. The possible responses are as follows: (1) At home; (2) In bars or restaurants; (3) At parties or events; (4) Others.**

To analyse this question, consumers who stated that they do not drink beer were excluded, resulting in a sample size of 921 respondents. This is a multiple-response question; thus, the sum of percentages exceeds 100 %. The results of the TMRS show that the majority of respondents consume beer in bars or restaurants (83.06 %), followed closely by consumption at home (77.2 %). A smaller proportion of participants reported consuming beer at parties or events (43.8 %), and only 0.98 % selected other places for beer consumption. Annexe B: Table B6 shows the distribution of the relative frequency of responses regarding places of beer consumption, including a comparison with the LLM-generated responses.

The first-order Rao-Scott corrected chi-square shows significant differences among the five response distributions (TMRS and four LLMs) with a value of 897.80, degrees of freedom (df) = 12, and p-value 0.0, leading to the rejection of the null hypothesis of distribution equality.

Complementary, the first four-order Rao-Scott corrected chi-square tests were performed to compare the response distributions of each LLM with the TMRS (Table 9). All the p-values were smaller than 0.01, leading to the rejection of all four null hypotheses. This means that none of the LLMs can fully replicate the response distribution of the TMRS regarding places of beer consumption, indicating significant differences.

Nevertheless, we observe differences in the performance of the AI-

based models. LLM\_2 and LLM\_3 have the smallest first-order Rao-Scott corrected chi-square, 50.57 and 78.14, respectively, indicating that their responses are more aligned with those of the TMRS compared to other LLMs.

This finding is also supported by the JSD and the MAD metrics, indicating high similarity for the LLM 2 (JSD = 0.0904 and MAD = 77.50) and LLM 3 (JSD = 0.1088 and MAD = 56.00), both with JSD and MAD lower than 0.2 and 100, respectively. This suggests they can serve as reliable proxies for estimating beer consumption in Spain. In contrast, the other LLMs show lower similarity, with JSD values greater than 0.2 but still below the cut-off of 0.5, indicating moderate alignment with the traditional survey responses and much higher MAD values (greater than 100).

Therefore, H1 is accepted, particularly for LLMs 2 and 3, as their responses, although they cannot fully replicate those provided by the TMRS, can serve as a proxy and complement them.

#### Summary of results: comparison of JSD in LLMs

Based on these results, Fig. 3 summarises the similarity of LLMs by question using JSD, which enables the comparison and ranking of the models' performance. LLM\_3 is the top performer in seven out of ten questions and ranks second in the remaining three, with an average JSD of 0.1412. This contrasts with the poorest performer, LLM\_2, which has an average JSD of 0.2295 (62.6 % higher).

Through our analysis, we have found that although LLMs cannot fully replicate TMRS, as all the chi-square tests and first-order Rao-Scott corrected chi-square tests led to the rejection of the null hypotheses of equality of distributions compared to the TMRS; they serve as effective proxies for the response distribution. The JSD of all the LLMs was <0.5 for all the questions, and for all but one question, there was always at least one model with a JSD <0.2, which represents a high similarity to

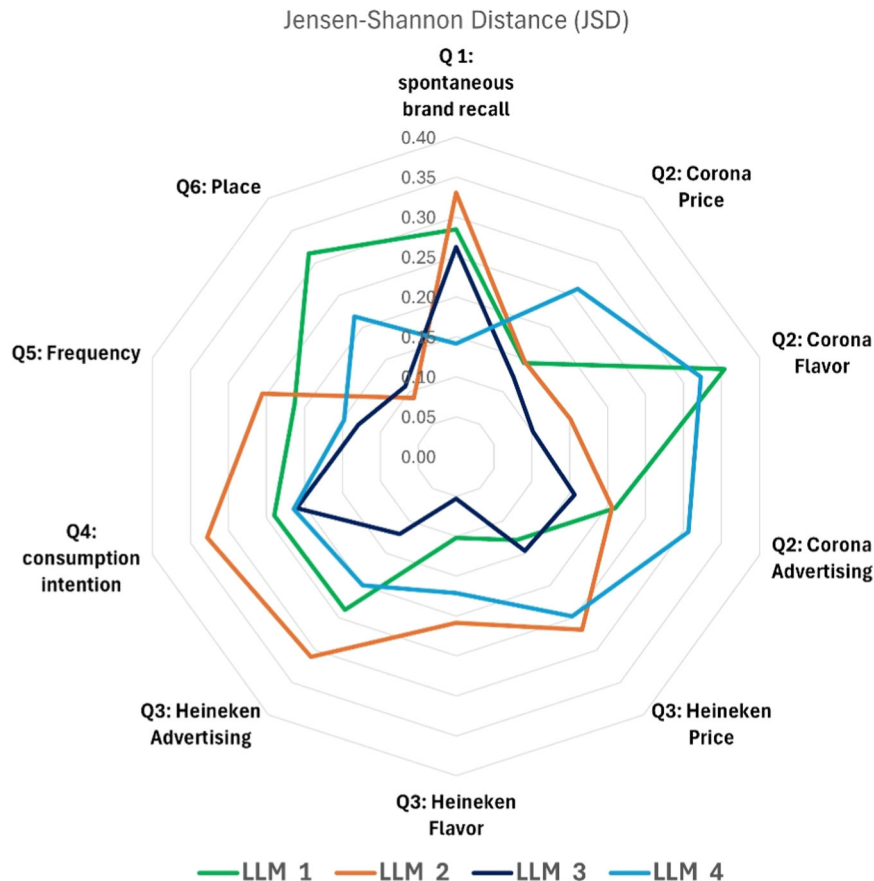


Fig. 3. Distribution similarity between TMRS and each LLM by question, measured using JSD.

TMRS.

However, Fig. 3 also highlights differences in the performance of the models. LLM\_3 has an average JSD of 0.1412, which is 36.69 %, 38.50 %, and 36.62 % lower than the averages for Models 1, 2, and 4, respectively. Additionally, the accuracy of the responses varies by both the model used and the type of question (Fig. 4), indicating that different models perform better or worse depending on the specific attributes being evaluated.

For example, JSD values range from a minimum of 0.0532 provided by LLM\_3 in Question 3 for evaluating the attribute ‘Flavour of Heineken’, to 0.3541 (565 % higher) provided by LLM\_1 in Question 2 for the evaluation of the attribute ‘Flavour of Corona’. This variability complicates understanding the errors made in the simulation when there is no real benchmark (TMRS) for comparison. Therefore, it is recommended to use several LLMs to corroborate and calibrate results.

These differences in the capabilities of the models to address different kinds of questions are summarised in Fig. 4. AI-based models are more effective at answering questions based on publicly available and observable data used to train them, such as consumption behaviours and product attributes, because of the abundance of information. This is evident in stages such as Consideration (Questions 2 and 3) and Experience (Questions 5 and 6) of the conversion funnel, which show very low average JSDs (0.1140 and 0.1097, respectively). However, they are less precise in predicting less documented aspects such as future purchase intentions or market share, apparent in stages such as Awareness (Question 1) and Purchase (Question 4), where higher JSDs (0.1415 and 0.2104 respectively) indicate that the lack of detailed data and the need to understand complex competitive and economic contexts increase the error compared to TMRS.

*Theoretical contributions and implications for practice*

None of the LLMs evaluated fully replicates the response distributions obtained through the traditional survey. Statistically significant differences were detected in all the questions analysed, confirming that these models, in their current state, are not perfect substitutes for human participation. However, and here lies one of the main theoretical contributions, similarity analysis using the JSD demonstrates that these models act as reliable and valuable proxies. The responses generated by LLMs show high similarity to human data, with JSD values mostly below 0.2. This finding validates our first hypothesis: LLMs can effectively complement traditional methods, especially for companies (including SMEs) seeking to conduct market research in a more agile and cost-effective way.

We also confirmed the hypothesis that LLMs can capture and reflect complex aspects of consumer psychology. The models could identify perceptions about brand attributes such as taste and advertising, and even reflected consumers’ tendency to be more critical of price. In addition, we provide empirical evidence for one of our novel hypotheses: the accuracy of LLMs decreases as the concreteness and specificity of the questions increase. The models are more effective at assessing

conversion funnel stages such as Consideration and Experience, which rely on more observable and publicly available data, but show greater discrepancies at the Brand Awareness (Top of Mind) and Purchase stages, where information is less accessible and the competitive context is more complex.

The implications for professional practice are profound. The main recommendation is to adopt a hybrid approach, where GAI does not replace but augments and enhances the capabilities of the human researcher. LLMs are excellent tools for exploratory phases, hypothesis generation, or scenario simulation. However, the remarkable variability in performance among the different models, with LLM\_3 being the most consistent in our analysis, underscores the critical need not to rely on a single tool. Practitioners must use multiple LLMs iteratively to corroborate and calibrate their findings, thereby mitigating the risk of errors in decision-making, especially when a benchmark such as a traditional survey is not available.

*Limitations and future research directions*

There are several restrictions on this study. Our TMRS focuses on the beer market in Spain; data from other industries should be examined in future studies.

In addition to this, to ensure a concise yet comprehensive analysis, we strategically selected six key questions that cover all stages of the conversion funnel. We aim to expand this research by replicating the analysis for the nine remaining questions and also by conducting an in-depth analysis for specific subgroups of respondents according to distinct customer segments.

**Conclusions**

AI—in particular, GAI—is evolving rapidly, and companies are eager to adopt it in various fields, especially in marketing, where it can provide a significant competitive advantage in an increasingly dynamic market. Our study contributes to the theoretical and empirical understanding of how GAI can reshape the foundations of market research and its implications for decision-making in companies.

Thus, we compare the capabilities of the four most popular LLMs available on the market (ChatGPT from OpenAI, Gemini from Google, Claude from Anthropic, and LLaMa from Meta) with TMR methods, such as surveys applied to the beer market in Spain. This comparison sheds light on both their benefits and drawbacks and provides practical recommendations for decision-makers on adopting these tools for performing market research studies. The mature and highly concentrated competitive beer market offers an ideal testing ground for comparing AI-based models and TMRS.

Our analysis shows that LLMs represent a valuable tool for performing market research. They can serve as a proxy to complement TMRS, responding with high similarity to most of the questions (JSD <0.2 in all questions but one, with a value of 0.2104), and are consistent with the distribution provided for the TMRS. Nevertheless, none of them

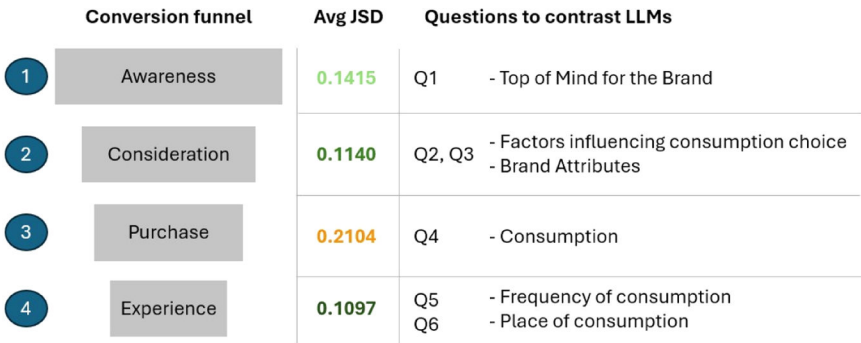


Fig. 4. Average JSD for responses generated by the best LLM at each stage of the conversion funnel.



can fully replicate the TMRS, confirming H1. Therefore, the adoption of LLMs can represent a competitive advantage not only for large companies, which can develop market research studies more efficiently through the combination of these models and TMRS, but also for SMEs that can now perform studies previously less accessible to them because of high costs.

Another relevant finding from this research is the confirmation of the hypothesis that LLMs can capture and reflect aspects of customer psychology, just as TMRS do. These aspects include personal preferences for product and brand attributes such as flavour and advertising, as well as the intrinsic tendency of customers to be more critical of product prices compared to other attributes. Compared to other types of questions, the average similarity for questions related to products and brand attributes is 20.16 % closer to the responses provided by the TMRS. This high similarity to the TMRS distribution responses demonstrated the capability of LLMs to effectively address questions related to customer psychology, confirming H2.

These findings also indicate variability in the performance among the models, with LLM\_3 emerging as the top performer in seven out of ten questions and ranking second in the remaining three with an average JSD of 0.1412. Additionally, the accuracy of the responses varies depending on both the model used and the type of question, showing a +565 % JSD difference from the most precise to the least precise response provided by an LLM. This indicates that different models perform better or worse depending on the specific aspects or attributes being evaluated. This variability complicates understanding the errors in the responses and represents a risk for companies in the decision-making process.

It stems from a combination of factors, including the training data and model architecture, temporal and contextual coverage, and the nature of the alignment mechanisms used during the development and training (also summarised in Table 1). In addition, the level of abstraction in the questions is also a very relevant factor, as the LLMs' accuracy decreases as questions become more concrete and specific. This is particularly evident in questions related to future purchase intentions, brand or product positioning, and market share, where highly detailed and context-specific information is required.

We found that the LLMs were more precise in funnel conversion stages related to Consideration (Question 2 about factors influencing consumption choice and Question 3 about brand attributes) and

Customer Experience (Question 5 about frequency of consumption and Question 6 about place of consumption), showing slight differences compared to the TMRS (JSD= 0.1140 and 0.1097 respectively), indicating minor discrepancies.

However, they are less accurate in predicting less documented aspects, such as future purchase intentions or market share, apparent in stages such as Awareness (Question 1 about top-of-mind brand recall) and Purchase Intentions (Question 4 about consumption) with JSD = 0.1415 and 0.2104, respectively.

These results show that the lack of detailed training data and the need to understand complex competitive and economic contexts to train the models impact the accuracy and relevance of their results compared to TMRS, thereby confirming H3.

Finally, another drawback is the reproducibility of results. The robustness of TMRS depends on the sample selection method, which aims to ensure the representativeness of the sample. In contrast, AI models are prone to producing varying results for the same questions. Practitioners and decision-makers must be aware of these risks and mitigate them by using multiple LLMs and employing an iterative process to corroborate and calibrate the results, especially when they do not have a TMRS to contrast these results.

Our findings indicate that LLMs can serve as proxies to replicate customer behaviour in market research studies, especially when these models are well-trained, calibrated, and benchmarked appropriately or when several LLMs are used simultaneously to triangulate results and enhance the capabilities of the human researcher. This challenges the boundaries between human and synthetic respondents and invites rethinking the theories related to customer behaviour, brand perception, and the value and efficiency that simulated data can bring to behavioural sciences.

#### CRedit authorship contribution statement

**Macarena Estevez:** Investigation, Formal analysis, Data curation, Conceptualization. **María Teresa Ballestar:** Writing – original draft, Methodology, Investigation, Formal analysis. **Jorge Sainz:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation.

#### Annexe A. Descriptive statistics of the sample under study in the TMRS.

**Table A1**  
Gender distribution.

Gender	Frequency	Percentage
Women	505	50.3 %
Men	498	49.7 %

**Table A2**  
Age distribution.

Age	Frequency	Percentage
18	8	0.9 %
19	16	1.7 %
20	11	1.2 %
21	9	1.0 %
22	13	1.4 %
23	17	1.8 %
24	25	2.7 %
25	15	1.6 %
26	18	1.9 %

(continued on next page)

**Table A2** (continued)

Age	Frequency	Percentage
27	12	1.3 %
28	11	1.2 %
29	17	1.8 %
30	16	1.7 %
31	17	1.8 %
32	17	1.8 %
33	16	1.7 %
34	24	2.6 %
35	24	2.6 %
36	18	1.9 %
37	21	2.2 %
38	24	2.6 %
39	21	2.2 %
40	14	1.5 %
41	28	3.0 %
42	15	1.6 %
43	14	1.5 %
44	18	1.9 %
45	25	2.7 %
46	19	2.0 %
47	23	2.5 %
48	18	1.9 %
51	22	2.4 %
52	19	2.0 %
53	20	2.1 %
54	14	1.5 %
55	14	1.5 %
56	21	2.2 %
57	23	2.5 %
58	24	2.6 %
59	18	1.9 %
60	24	2.6 %
61	21	2.2 %
62	19	2.0 %
63	11	1.2 %
64	5	0.5 %
65	28	3.0 %
66	18	1.9 %
67	20	2.1 %
68	14	1.5 %
69	13	1.4 %
70	8	0.9 %
71	7	0.7 %
72	10	1.1 %
73	10	1.1 %
74	4	0.4 %
75 or more	3	0.3 %

**Table A3**  
Social Class distribution.

Social Class	Frequency	Percentage
High	13	1.3 %
Upper middle	90	9.0 %
Middle	602	60.0 %
Lower middle	240	23.9 %
Low	58	5.8 %

**Table A4**  
Occupation distribution.

Occupation	Frequency	Percentage
Employed (salaried worker)	608	60.6 %
Retired	132	13.2 %
Self-employed (freelancer)	84	8.4 %
Job seeker	81	8.1 %
Studying	55	5.5 %
Domestic work	43	4.3 %

**Table A5**  
Household Composition distribution.

Currently living	Frequency	Percentage
With my partner and my children	395	39.4 %
With my partner	255	25.4 %
With other family members	152	15.2 %
Alone	127	12.7 %
With my children	57	5.7 %
With other people	17	1.7 %

**Table A6**  
Education distribution.

Education	Frequency	Percentage
No education	4	0.4 %
Basic education	125	12.5 %
Intermediate education	374	37.3 %
Higher education	500	49.9 %

**Table A7**  
Income distribution.

Income	Frequency	Percentage
Less than €2000	345	34.4 %
From €2000 to €3000	346	34.5 %
From €3000 to €4000	194	19.3 %
More than €4000	118	11.8 %

**Table A8**  
Household Children Composition.

Children	Frequency	Percentage
I have no children	388	33.9 %
Under 5 years old	97	8.5 %
From 5 to 9 years old	115	10.0 %
From 10 to 14 years old	156	13.6 %
From 15 to 19 years old	141	12.3 %
20 years old or older	249	21.7 %

**Table A9**  
Location distribution.

Location	Frequency	Percentage
Madrid	89	8.9 %
Barcelona	62	6.2 %
Sevilla	30	3.0 %
Zaragoza	24	2.4 %
Valencia/València	22	2.2 %
A Coruña	15	1.5 %
Alicante/Alacant	15	1.5 %
Murcia	15	1.5 %
Bizkaia	12	1.2 %
Las Palmas	12	1.2 %
Málaga	12	1.2 %
Burgos	11	1.1 %
Cádiz	9	0.9 %
Granada	9	0.9 %
Valladolid	9	0.9 %
Barcelona - Badalona	8	0.8 %
...	...	...

# Annexe B. Distribution of relative frequencies for beer survey questions: TMRS vs. LLMs

Table B1

Question 1: Distribution of the relative frequency of responses regarding top-of-mind beer brands among customers.

Model Type	Mahou	Estrella Galicia	Heineken	Cruzcampo	Estrella Damm	San Miguel	Alhambra	Amstel	Others	TOTAL
TMRS	24.53 %	13.86 %	10.57 %	8.47 %	7.68 %	5.48 %	4.69 %	4.19 %	20.54 %	100 %
LLM_1	15.00 %	38.00 %	11.00 %	10.00 %	6.00 %	7.00 %	0.00 %	8.00 %	5.00 %	100 %
LLM_2	33.60 %	0.00 %	5.30 %	11.90 %	22.70 %	3.00 %	14.10 %	2.00 %	7.40 %	100 %
LLM_3	18.30 %	12.80 %	7.00 %	16.10 %	16.60 %	15.00 %	4.00 %	7.20 %	3.00 %	100 %
LLM_4	19.00 %	5.00 %	11.00 %	9.00 %	14.00 %	8.00 %	7.00 %	4.00 %	23.00 %	100 %

Table B2

Question 2: Distribution of the relative frequency of responses regarding customers' perceived value of specific beer attributes for the CORONA BRAND.

Corona: Price Attribute						
Model Type	Very_good	Good	Bad	Very_bad	DK/NA	TOTAL
TMRS	11.36 %	45.20 %	26.57 %	4.19 %	12.68 %	100 %
LLM_1	12.90 %	38.04 %	31.20 %	12.46 %	5.40 %	100 %
LLM_2	15.10 %	31.42 %	21.50 %	10.25 %	21.72 %	100 %
LLM_3	13.12 %	47.08 %	24.26 %	10.25 %	5.29 %	100 %
LLM_4	17.53 %	20.95 %	17.53 %	19.63 %	24.37 %	100 %
Corona: Flavour Attribute						
Model Type	Very_good	Good	Bad	Very_bad	DK/NA	TOTAL
TMRS	21.17 %	53.25 %	11.58 %	3.64 %	10.36 %	100 %
LLM_1	10.69 %	22.93 %	41.68 %	20.73 %	3.97 %	100 %
LLM_2	20.51 %	35.17 %	18.30 %	8.49 %	17.53 %	100 %
LLM_3	27.12 %	45.42 %	15.99 %	5.95 %	5.51 %	100 %
LLM_4	14.77 %	17.97 %	23.48 %	20.62 %	23.15 %	100 %
Corona: Advertising Attribute						
Model Type	Very_good	Good	Bad	Very_bad	DK/NA	TOTAL
TMRS	16.87 %	48.73 %	14.66 %	2.09 %	17.64 %	100 %
LLM_1	19.29 %	40.13 %	21.72 %	13.12 %	5.73 %	100 %
LLM_2	12.13 %	28.45 %	23.15 %	12.13 %	24.15 %	100 %
LLM_3	23.59 %	51.05 %	14.88 %	5.29 %	5.18 %	100 %
LLM_4	19.63 %	18.08 %	18.41 %	24.48 %	19.40 %	100 %

Table B3

Question 3: Distribution of the relative frequency of responses regarding customers' perceived value of specific beer attributes for the HEINEKEN BRAND.

Heineken: Price Attribute					
Very_good	Good	Bad	Very_bad	DK/NA	TOTAL
14.32 %	53.32 %	20.23 %	3.01 %	9.13 %	100 %
12.45 %	51.87 %	25.93 %	7.26 %	2.49 %	100 %
10.37 %	22.82 %	29.05 %	12.45 %	25.31 %	100 %
9.85 %	44.29 %	29.25 %	10.68 %	5.91 %	100 %
30.29 %	27.28 %	12.14 %	7.57 %	22.72 %	100 %
Heineken: Flavour Attribute					
Very_good	Good	Bad	Very_bad	DK/NA	TOTAL
25.31 %	51.87 %	11.00 %	4.46 %	7.37 %	100 %
31.12 %	46.68 %	12.45 %	7.26 %	2.49 %	100 %
18.67 %	31.12 %	20.75 %	8.30 %	21.16 %	100 %
24.79 %	49.79 %	15.77 %	3.84 %	5.81 %	100 %
39.63 %	31.74 %	7.99 %	4.77 %	15.87 %	100 %
Heineken: Advertising Attribute					
Very_good	Good	Bad	Very_bad	DK/NA	TOTAL
31.12 %	51.35 %	5.81 %	2.07 %	9.65 %	100 %
15.56 %	57.05 %	18.67 %	7.26 %	1.45 %	100 %
14.52 %	26.97 %	24.90 %	10.37 %	23.24 %	100 %
30.71 %	46.58 %	13.38 %	4.46 %	4.88 %	100 %
33.30 %	27.80 %	12.97 %	7.37 %	18.57 %	100 %



**Table B4**  
Question 4: Distribution of the relative frequency of responses regarding customers' brand consumption intention.

Model Type	Mahou	Estrella Galicia	Heineken	Cruzcampo	Estrella Damm	San Miguel	Alhambra	Amstel	Águila	Corona	Others	TOTAL
TMRS	42.35 %	49.84 %	30.62 %	23.45 %	22.69 %	25.41 %	27.14 %	15.96 %	15.20 %	15.20 %	41.04 %	308.90 %
LLM_1	40.17 %	43.11 %	11.83 %	8.03 %	6.19 %	5.21 %	3.80 %	9.88 %	0.00 %	3.15 %	26.71 %	158.09 %
LLM_2	38.00 %	25.95 %	14.01 %	9.01 %	3.04 %	2.50 %	1.95 %	0.76 %	1.52 %	0.33 %	2.61 %	99.67 %
LLM_3	45.17 %	35.83 %	35.61 %	33.44 %	31.16 %	19.22 %	12.05 %	9.45 %	7.17 %	4.78 %	4.78 %	238.65 %
LLM_4	23.56 %	16.50 %	14.88 %	0.65 %	10.97 %	8.03 %	6.30 %	6.30 %	0.65 %	0.98 %	19.76 %	108.58 %

**Table B5**  
Question 5: Distribution of the relative frequency of responses regarding beer consumption frequency by model type.

Model Type	Daily	Once_week	Once_month	Less_Frequency	Never	TOTAL
TMRS	24.90 %	48.90 %	10.50 %	7.60 %	8.20 %	100 %
LLM_1	16.32 %	29.33 %	23.92 %	20.02 %	10.41 %	100 %
LLM_2	15.00 %	25.00 %	30.00 %	20.00 %	10.00 %	100 %
LLM_3	15.00 %	45.00 %	20.00 %	12.00 %	8.00 %	100 %
LLM_4	15.00 %	40.00 %	20.00 %	10.00 %	15.00 %	100 %

**Table B6**  
Question 6: Distribution of the relative frequency of responses regarding places of beer consumption.

Model Type	In bars or restaurants	At home	At parties or events	Others	TOTAL
TMRS	83.06 %	77.20 %	43.76 %	0.98 %	204.99 %
LLM_1	31.05 %	31.16 %	30.18 %	33.55 %	125.95 %
LLM_2	59.72 %	48.86 %	27.14 %	5.43 %	141.15 %
LLM_3	85.00 %	65.00 %	45.00 %	10.00 %	205.00 %
LLM_4	93.70 %	82.52 %	64.60 %	41.80 %	282.63 %

References

Abumalloh, R., Niasah, M., & Ali, R. (2024). Does metaverse improve communications quality and customer trust? A user-centric evaluation framework based on the cognitive-affective-behavioural theory. *Journal of Innovation & Knowledge*, 9(1), Article 100569. <https://doi.org/10.1016/j.jik.2024.100569>

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S., & Anadkat, S. (2023). Gpt-4 technical report. *ArXiv Preprint ArXiv:2303.08774*.

Anil, R., Dai, A. M., Firat, O., Johnson, M., Lepikhin, D., Passos, A., & Wu, Y. (2023). Gemini: A family of highly capable multimodal models BT - Google AI blog. <http://blog.google/intl/es-es/productos/tecnologia/presentamos-gemini-nuestro-mo-del-del-ia/>.

Anthis, J. R., Liu, R., Richardson, S. M., Kozlowski, A. C., Koch, B., Evans, J., Brynjolfsson, E., & Bernstein, M. (2025). LLM social simulations are a promising research method. <http://arxiv.org/abs/2504.02234>.

Arora, N., Chakraborty, I., & Nishimura, Y. (2024). EXPRESS: AI-Human hybrids for marketing research: Leveraging LLMs as collaborators. *Journal of Marketing*. <https://doi.org/10.1177/00222429241276529>

Atari, M., Xue, Mona, J., Park, P.S., Blasi, D.E., & Henrich, J. (2023). Which humans? *PsyArXiv Preprint*, 1–23. <https://doi.org/10.31234/osf.io/5b26t>.

Bai, Y., Jones, A., Ndousse, K., Askell, A., Chen, A., DasSarma, N., Drain, D., Fort, S., Ganguli, D., & Henighan, T. (2022). Training a helpful and harmless assistant with reinforcement learning from human feedback. *ArXiv Preprint ArXiv:2204.05862*.

Bilder, C. R., & Loughin, T. M. (2001). On the first-order Rao–Scott correction of the Umes–Loughin–Scherer statistic. *Biometrics*, 57(4), 1253–1255. <https://doi.org/10.1111/j.0006-341X.2001.01253.x>

Blair, C., Weigel, D. E., Balazik, M., Keeley, A. T. H., Walker, F. M., Landguth, E., Cushman, S., Murphy, M., Waits, L., & Balkenhol, N. (2012). A simulation-based evaluation of methods for inferring linear barriers to gene flow. *Molecular Ecology Resources*, 12(5), 822–833. <https://doi.org/10.1111/j.1755-0998.2012.03151.x>

Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Advances in Neural Information Processing Systems*, 29.

Brand, J., Israeli, A., & Ngwe, D. (2023). Using LLMs for Market Research. Harvard Business School Marketing Unit Working Paper No. 23-062, Available at SSRN: <http://ssrn.com/abstract=4395751> or <https://doi.org/10.2139/ssrn.4395751>.

Brand, J., Israeli, A., & Ngwe, D. (2024). *Using LLMs for market research* (HBS WP).

Briët, J., & Harremoës, P. (2009). Properties of classical and quantum Jensen-Shannon divergence. *Physical Review A*, 79(5), Article 052311. <https://doi.org/10.1103/PhysRevA.79.052311>

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

Campbell, C., Sands, S., McFerran, B., & Mavrommatis, A. (2023). Diversity representation in advertising. *Journal of the Academy of Marketing Science*. <https://doi.org/10.1007/s11747-023-00994-8>

Cano-Marín, E. (2024). The transformative potential of Generative Artificial Intelligence (GenAI) in business: A textmining analysis on innovation data sources. *ESIC Market. Economics & Business Journal*, 55(2), e333. <https://doi.org/10.7200/esicm.55.333>

Cao, X., & Kosinski, M. (2024). Large language models know how the personality of public figures is perceived by the general public. *Scientific Reports*, 14(1). <https://doi.org/10.1038/s41598-024-57271-z>

Chen, J. W., Bersamin, M., Waiters, E., Keefe, D. B., & G, M. J. (2005). Alcohol advertising: What makes it attractive to youth? *Journal of Health Communication*, 10 (6), 553–565. <https://doi.org/10.1080/10810730500228904>

Chiarello, F., Giordano, V., Spada, I., Barandoni, S., & Fantoni, G. (2024). Future applications of generative large language models: A data-driven case study on ChatGPT. *Technovation*, 133. <https://doi.org/10.1016/j.technovation.2024.103002>

Cillo, P., & Rubera, G. (2024). Generative AI in innovation and marketing processes: A roadmap of research opportunities. *Journal of the Academy of Marketing Science*. <https://doi.org/10.1007/s11747-024-01044-7>

Colicev, A., Kumar, A., & O'Connor, P. (2019). Modeling the relationship between firm and user generated content and the stages of the marketing funnel. *International Journal of Research in Marketing*, 36(1), 100–116.

Connor, R., Cardillo, F. A., Moss, R., & Rabitti, F. (2013). Evaluation of Jensen-Shannon distance over sparse data. In *Similarity Search and Applications: 6th International Conference, SISAP 2013, A Coruña, Spain, October 2-4, 2013, Proceedings* (pp. 163–168). Springer Berlin Heidelberg.

- Costa, R., & Di Pillo, F. (2024). Aligning innovative banks' sustainability strategies with customer expectations and perceptions: The CSR feedback framework. *Journal of Innovation & Knowledge*, 9(4), Article 100596. <https://doi.org/10.1016/j.jik.2024.100596>
- Demszyk, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., & Ravi, S. (2020). GoEmotions: A dataset of fine-grained emotions. <http://arxiv.org/abs/2005.00547>.
- Devlin, J. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *ArXiv Preprint ArXiv:1810.04805*.
- Dwivedi, Y. K. (2025). Generative Artificial Intelligence (GenAI) in entrepreneurial education and practice: Emerging insights, the GAIN framework, and research agenda. *International Entrepreneurship and Management Journal*, 21, 82. <https://doi.org/10.1007/s11365-025-01089-2>
- Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). GPTs are GPTs: An early look at the labor market impact potential of large language models. <http://arxiv.org/abs/2303.10130>.
- Estevez, M., Ballestar, M. T., & Sainz, J. (2024). A primer on out-of-the-box AI marketing mix models. *IEEE Transactions on Engineering Management*.
- Garson, G. I., & Moser, E. B. (1995). Aggregation and the Pearson chi-square statistic for homogeneous proportions and distributions in ecology. *Ecology*, 76(7), 2258–2269. <https://doi.org/10.2307/1941693>
- GonzálezPadilla, P., HernándezTamurejo, Á., & Debasa, F. (2025). The dissociation of user identity and behaviour in the real world versus the virtual world: A review. *ESIC Market. Economics & Business Journal*, 56(1), e399. <https://doi.org/10.7200/esicm.56.399>
- Goretti, L., Küsters, A., Morgana, S., Sharma, A., Goretti, L., Küsters, A., Morgana, S., Dec, A. S., & Goretti, L. (2024). Challenges and opportunities of using generative AI for research : Opening the discussion Challenges and opportunities of using generative AI for research: Opening the discussion. *The International Spectator*, 0(0), 1–16. <https://doi.org/10.1080/03932729.2024.2435176>
- Green, P. E., & Srinivasan, V. (1978). Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research*, 5(2), 103–123. <https://doi.org/10.1086/208721>
- Grewal, D., Guha, A., Satornino, C. B., & Becker, M. (2024a). The future of marketing and marketing education. *Journal of Marketing Education*. <https://doi.org/10.1177/02734753241269838>
- Grewal, D., Satornino, C. B., Davenport, T., & Guha, A. (2024b). How generative AI is shaping the future of marketing. *Journal of The Academy of Marketing Science*. <https://doi.org/10.1007/s11747-024-01064-3>
- Heo, S. (2006). Power analysis of the Rao-Scott first-order adjustment to the Pearson test for homogeneity. In *Joint Statistical Meetings Proceedings* (pp. 3126–3129).
- Hermann, E., & Puntoni, S. (2024). Artificial intelligence and consumer behavior: From predictive to generative AI. *Journal of Business Research*, 180. <https://doi.org/10.1016/j.jbusres.2024.114720>
- JuárezMadrid, D., & Monreal, E. G. (2024). Marketoriented entrepreneurship and the impact of social media and crowdsourcing on continuous organizational learning: An empirical study. *Sustainable Technology and Entrepreneurship*, 4(2), Article 100088. <https://doi.org/10.1016/j.jstae.2024.100088>
- Kirk, P., & Givi, J. (2025). The AI-authorship effect: Understanding authenticity, moral disgust, and consumer responses to AI-generated marketing communications. *Journal of Business Research*, 186. <https://doi.org/10.1016/j.jbusres.2024.114984>
- Kshetri, N. (2023). Generative artificial intelligence in marketing. *IT Professional*, 25(5), 71–75. <https://doi.org/10.1109/MITP.2023.3314325>
- Kshetri, N. (2024). Generative AI in advertising. *IT Professional*, 26(5), 15–19. <https://doi.org/10.1109/MITP.2024.3457328>
- Kshetri, N., Dwivedi, Y. K., Davenport, T. H., & Panteli, N. (2024). Generative artificial intelligence in marketing: Applications, opportunities, challenges, and research agenda. *International Journal of Information Management*, 75. <https://doi.org/10.1016/j.ijinfomgt.2023.102716>
- Kumar, V., Kotler, P., Gupta, S., & Rajan, B. (2024). Generative AI in marketing: Promises, perils, and public policy implications. *Journal of Public Policy & Marketing*. <https://doi.org/10.1177/07439156241286499>
- Lavassani, K. M., Movahedi, B., & Kumar, V. (2009). Developments in analysis of multiple response survey data in categorical data analysis: The case of enterprise system implementation in large North American firms. *Journal of Applied Quantitative Methods*, 4(1). [https://jaqm.rn/issuues/volume-4,issue-1/pdfs/lavassani\\_movahedi\\_kumar.pdf](https://jaqm.rn/issuues/volume-4,issue-1/pdfs/lavassani_movahedi_kumar.pdf)
- Lee, G., & Kim, H. Y. (2024). Algorithm fashion designer? Ascribed mind and perceived design expertise of AI versus human. *Psychology and Marketing*. <https://doi.org/10.1002/mar.22124>
- Lee, G., Lee, K., Jeong, B., & Kim, T. (2024). Developing Personalized Marketing Service Using Generative AI. *IEEE Access*, 12, 22394–22402. <https://doi.org/10.1109/ACCESS.2024.3361946>
- Lerro, G., Nazzaro, C., & M. M. (2020). Measuring consumers' preferences for craft beer attributes through best-worst scaling. *Agricultural and Food Economics*, 8(1), 1. <https://doi.org/10.1186/s40100-019-0147-4>
- Li, P., Castelo, N., Katona, Z., & Sarvary, M. (2024). Frontiers: Determining the validity of large language models for automated perceptual analysis. *Marketing Science*, 43(2), 254–266. <https://doi.org/10.1287/mksc.2023.0454>
- Malone, T., & Lusk, J. L. (2018). If you brew it, who will come? Market segments in the U.S. beer market. *Agribusiness*, 34(2), 204–221. <https://doi.org/10.1002/agr.21511>
- MarketsandMarkets. (2024). Digital transformation market size, share & trends. In *MarketsandMarkets*.
- Matz, S. C., Teeny, J. D., Vaid, S. S., Peters, H., Harari, G. M., & Cerf, M. (2024). The potential of generative AI for personalized persuasion at scale. *Scientific Reports*, 14 (1). <https://doi.org/10.1038/s41598-024-53755-0>
- Meyerding, A., & Lehberger, M. (2019). Consumer preferences for beer attributes in Germany: A conjoint and latent class approach. *Journal of Retailing and Consumer Services*, 47, 229–240. <https://doi.org/10.1016/j.jretconser.2018.12.007>
- Mikolov, T. (2013). Efficient estimation of word representations in vector space. *ArXiv Preprint ArXiv:1301.3781*, 3781.
- Nave, E., Duarte, P., Rodrigues, R. G., Paço, A., Alves, H., & Oliveira, T. (2022). Craft beer – a systematic literature review and research agenda. *International Journal of Wine Business Research*, 34(2), 278–307. <https://doi.org/10.1108/IJWBR-05-2021-0029>
- Niculescu, L., & Rîpă, I. A. (2024). Linking innovative work behavior with customer experience and marketing performance. *Journal of Innovation & Knowledge*, 9(2), Article 100560. <https://doi.org/10.1016/j.jik.2024.100560>
- Noel, T. F., Grady, J. J., & B. J. K. (2018). Advertising content, platform characteristics and the appeal of beer advertising on a social networking site. *Alcohol and Alcoholism*, 53(5), 619–625. <https://doi.org/10.1093/alcalc/agy040>
- Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654), 187–192. <https://doi.org/10.1126/science.adh2586>
- OpenAI. (2023). GPT-4 technical report. *ArXiv Preprint*. <https://doi.org/10.48550/arXiv.2303.08774>
- Orth, K., & L. U. R. (2006). Consumer-based brand equity versus product-attribute utility: A comparative approach for craft beer. *Journal of Food Products Marketing*, 11(4), 77–90. [https://doi.org/10.1300/J038v11n04\\_06](https://doi.org/10.1300/J038v11n04_06)
- Pagani, M., & Wind, Y. (2024). Unlocking marketing creativity using artificial intelligence. *Journal of Interactive Marketing*. <https://doi.org/10.1177/10949968241265855>
- Park, J., & Ahn, S. (2024). Traditional vs. AI-generated brand personalities: Impact on brand preference and purchase intention. *Journal of Retailing and Consumer Services*, 81. <https://doi.org/10.1016/j.jretconser.2024.104009>
- Park, J. S., Popowski, L., Cai, C., Morris, M. R., Liang, P., & Bernstein, M. S. (2022). Social Simulacra: Creating populated prototypes for social computing systems. In *UIST 2022 - Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. <https://doi.org/10.1145/3526113.3545616>
- Peng, S. (2023). Sharing economy and sustainable supply chain perspective: The relation of sustainable, economic and social pillar of supply chain in customer retention and sustainable development. *Journal of Innovation & Knowledge*, 8(1), Article 100316. <https://doi.org/10.1016/j.jik.2023.100316>
- Peres, R., Schreier, M., Schweidel, D., & Sorescu, A. (2023). On ChatGPT and beyond: How generative artificial intelligence may affect research, teaching, and practice. *International Journal of Research in Marketing*, 40(2), 269–275. <https://doi.org/10.1016/j.ijresmar.2023.03.001>
- Pinto, S., Albanese, F., Dorso, C. O., & Balenzuela, P. (2018). A quantitative analysis of Media Agenda and Public Opinion using time-evolving topic distribution. *ArXiv Preprint*. <https://doi.org/10.48550/arXiv.1807.05184>
- PradosCastillo, J. F., TorrecillaGarcía, J. A., GuaitaFernández, P., & De CastroPardo, M. (2024). The impact of the metaverse on consumer behaviour and marketing strategies in tourism: A bibliometric review. *ESIC Market. Economics & Business Journal*, 55(1), e327. <https://doi.org/10.7200/esicm.55.327>
- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140), 1–67.
- Ricciuto, L., Tarasuk, V., & Yatchew, A. (2006). Socio-demographic influences on food purchasing among Canadian households. *European Journal of Clinical Nutrition*, 60 (6), 778–790. <https://doi.org/10.1038/sj.ejcn.1602382>
- Rostamzadeh, R., Bakayono, M., Strielkowski, W., & Streimikiene, D. (2024). Providing an improved method for social customer relationship management: Meta analysis approach. *Journal of Innovation & Knowledge*, 9(3), Article 100566. <https://doi.org/10.1016/j.jik.2024.100566>
- Sadrani, R., Sharma, H., & Bachan, A. (2024). On the relationship between visual anomaly-free and anomalous representations. *ArXiv Preprint*. <https://doi.org/10.48550/arXiv.2410.06576>
- Sarstedt, M., Adler, S. J., Rau, L., & Schmitt, B. (2024). Using large language models to generate silicon samples in consumer and marketing research: Challenges, opportunities, and guidelines. *Psychology and Marketing*, 41(6), 1254–1270. <https://doi.org/10.1002/mar.21982>
- Saura, J. R., & Ruizskiene, R. (2025). Behavioral economics, artificial intelligence and entrepreneurship: An updated framework for management. *International Entrepreneurship and Management Journal*, 21, 67. <https://doi.org/10.1007/s11365-025-01076-7>
- Singh, S. K., Kumar, S., & Mehra, P. S. (2023, June). Chat gpt & google bard ai: A review. In *2023 International Conference on IoT, Communication and Automation Technology (ICIAT)* (pp. 1–6). IEEE. <https://doi.org/10.1109/ICIAT57735.2023.10263706>
- Song, M. H., Kang, E. S., Jeong, C. H., Chow, M. Y., & Ayhan, B. (2003). Mean absolute difference approach for induction motor broken rotor bar fault detection BT -. In *4th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives*, 2003. SDEMPED 2003 (pp. 115–118). IEEE. <https://doi.org/10.1109/SDEMPED.2003.1234557>
- Spinner, M. A., Schaffert, S. A., Stehr, H., Santaguida, M. T., Kita, R., Aleshin, A., Zehnder, J. L., & Greenberg, P. L. (2021). MDS and MDS/MPN genomic subgroups demonstrate differential E x vivo drug sensitivity. *Blood*, 138, 4645.
- Szczesniowski, J. J., Alba, A. R., Castro, P. R., Gómez, M. L., González, J. S., & González, L. L. (2024). Quality of information about urologic pathology in English and Spanish from ChatGPT, BARD, and Copilot. *Actas Urológicas Españolas (English Edition)*, 48(5), 398–403. <https://doi.org/10.1016/j.acuroe.2024.02.009>

- Tafesse, W., & Wien, A. (2024). ChatGPT's applications in marketing: A topic modeling approach. *Marketing Intelligence & Planning*, 42(4), 666–683. <https://doi.org/10.1108/MIP-10-2023-0526>
- Thomé, K., Soares, A. P., & Moura, J. V. (2017). Social interaction and beer consumption. *Journal of Food Products Marketing*, 23(2), 186–208. <https://doi.org/10.1080/10454446.2017.1244797>
- Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). *LLaMA: Open and efficient foundation language models*. *ArXiv Preprint* <https://doi.org/10.48550/arXiv.2302.13971>.
- Vrontis, D. (1998). Strategic assessment: The importance of branding in the European beer market. *British Food Journal*, 100(2), 76–84. <https://doi.org/10.1108/00070709810204066>
- Wach, K., Duong, C. D., Ejidys, J., Kazlauskaitė, R., Korzynski, P., Mazurek, G., Paliszkiwicz, J., & Ziemba, E. (2023). The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. *Entrepreneurial Business and Economics Review*, 11(2), 7–30. <https://doi.org/10.15678/EBER.2023.110201>
- Wahid, R., Mero, J., & Ritala, P. (2023). Written by ChatGPT, illustrated by Midjourney: Generative AI for content marketing. *Asia Pacific Journal of Marketing and Logistics*, 35(8), 1813–1822. <https://doi.org/10.1108/APJML-10-2023-994>
- Wang, F., Yan, Z., Luan, Y., & Zhang, H. (2023). Blockchain adoption and security management of large scale industrial renewable resources: Knowledge management dimension. *Journal of Innovation & Knowledge*, 8(1), Article 100328. <https://doi.org/10.1016/j.jik.2023.100328>
- Wang, H., & Lu, W. C. (2025). AI-induced job impact: Complementary or substitution? *Sustainable Technology and Entrepreneurship*, 4(1), Article 100085. <https://doi.org/10.1016/j.stae.2024.100085>
- Ye, Y., Tong, C., Dong, B., Huang, C., Bao, H., & Deng, J. (2024). Alleviate light pollution by recognizing urban night-time light control area based on computer vision techniques and remote sensing imagery. *Ecological Indicators*, 158, Article 111591. <https://doi.org/10.1016/j.ecolind.2023.111591>
- Yenduri, G., Ramalingam, M., Selvi, G. C., Supriya, Y., Srivastava, G., Maddikunta, P. K. R., Raj, G. D., Jhaveri, R. H., Prabadevi, B., Wang, W., Vasilakos, A. V., & Gadekallu, T. R. (2024). GPT (Generative Pre-Trained Transformer) - A comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. *IEEE Access*, 12(March), 54608–54649. <https://doi.org/10.1109/ACCESS.2024.3389497>
- Zhou, W., Zhang, C., Wu, L., & Shashidhar, M. (2023). ChatGPT and marketing: Analyzing public discourse in early twitter posts. *Journal of Marketing Analytics*, 11(4), 693–706. <https://doi.org/10.1057/s41270-023-00250-6>
- Zhu, Y., & Zhang, Y. (2023). How to balance the dual functions of resources and capabilities to improve customer performance? *Journal of Innovation & Knowledge*, 8(1), Article 100311. <https://doi.org/10.1016/j.jik.2023.100311>