





Using LLMs in sensory service research: initial insights and perspectives

在感官服务研究中使用大语言模型：初步见解与展望

Monika Imschloss ^a, Marko Sarstedt ^{b,c}, Susanne J. Adler ^b and Jun Hwa Cheah ^d

^aInstitute of Management & Organization, Leuphana University of Lüneburg Lüneburg, Germany; ^bInstitute for Marketing, LMU Munich School of Management, Ludwig-Maximilians-University Munich, Munich, Germany; ^cFaculty of Economics and Business Administration, Babeş-Bolyai University, Cluj-Napoca, Romania; ^dNorwich Business School, University of East Anglia, Norwich, UK

ABSTRACT

Researchers have started using large language models (LLMs), such as OpenAI's GPT, to generate synthetic datasets designed to mimic human response behavior. Several studies have systematically compared LLM-generated data with human samples in order to explore LLMs' ability to mimic consumer decision-making. Extending prior findings, our research sets out to explore how GPT-4o responds to sensory information, and to evaluate its ability to grasp crossmodal correspondences as well as multisensory congruence – as commonly encountered in service settings. Our results indicate that while GPT-4o identifies and describes sensory stimuli accurately, it often fails to replicate the associative meanings and interpretations that humans derive from these stimuli, especially in stand-alone assessments. Our research therefore underscores the need for further exploration of the conditions under which LLMs reliably mirror human responses to sensory stimuli, and the implications of using LLMs in research on sensory-rich service settings.

摘要

研究人员已经开始使用大语言模型 (LLMs), 例如OpenAI的GPT, 来生成旨在模拟人类响应行为的合成数据集。一些研究系统地将大语言模型生成的数据与人类样本进行了比较, 以探索大语言模型模仿消费者决策的能力。延续先前的研究, 我们的研究旨在探索GPT-4o如何响应感官信息, 并评估其理解跨模态对应关系以及多感官一致性的能力, 而这在服务场景中是非常常见的。我们的研究表明, 尽管GPT-4o能够准确识别和描述感官刺激, 但它往往无法复制人类从这些刺激中得出的联想意义和解释, 尤其是在独立评估中。因此, 我们的研究强调了进一步探索大语言模型在何种条件下能够可靠地反映人类对感官刺激的反应, 以及在感官丰富的服务场景研究中使用大语言模型的意义。

ARTICLE HISTORY



Received 2 October 2024
Accepted 11 March 2025

KEYWORDS

Generative artificial intelligence; large language models; sensory marketing; service research; servicescape

关键词

生成式人工智能; 大语言模型; 感官营销; 服务研究; 服务场景。

CONTACT Marko Sarstedt  sarstedt@lmu.de  Institute for Marketing, LMU Munich School of Management, Ludwig-Maximilians-University Munich, Ludwigstr. 28RG, 80539 Munich, Germany

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

1. Introduction

Generative artificial intelligence (AI) and, particularly, large language models (LLMs) like OpenAI's Generative Pre-trained Transformer (GPT) are transforming marketing research as we know it (Hermann & Puntoni, 2024; Sarstedt, Adler, et al., 2024). As a computational simulation of human speech that potentially also inherits covert social information, GPT has the potential to serve as 'an invaluable source of insight into consumer preferences due to its ability to mimic or replicate human responses' (Brand et al., 2023, p. 2). Researchers have recently started testing this notion by systematically comparing synthetic (or silicon) samples with human samples to explore LLMs' ability to adequately mimic behavioral phenomena. Corresponding studies spanning fields as diverse as human-computer interaction, general psychology, and social psychology, produced mixed results in this regard. While GPT has, for example, been found to be sensitive to various types of biases regarding judgment (e.g. probability weighting), information processing (e.g. framing effects), and response behavior (e.g. social desirability bias) as well as decision-making heuristics (e.g. confirmation bias) (e.g. Binz & Schulz, 2023; Gao et al., 2024; Salecha et al., 2024), researchers could not identify other effects that are known to characterize consumer behavior, such as the endowment effect, mental accounting, and prospect theory elements (Y. Chen et al., 2023). Summarizing the diverse research in this field, synthetic participants might indeed, under specific conditions, be adequate proxies for human participants (e.g. Arora et al., 2025; Demszky et al., 2023; Sarstedt, Adler, et al., 2024), thereby offering both service researchers and companies enormous potentials.

While research on LLM-human comparisons is evolving very quickly, the applicability of LLMs in sensory consumer research is a field that has remained largely unexplored; that is, can LLMs, like GPT, be used to mimic sensory stimuli's impact on consumer behavior? This is surprising since subtle, and often unconsciously perceived sensory stimuli are known to shape consumer perceptions of and responses to service environments (e.g. Ahn & Back, 2020; Girard et al., 2019; Sarstedt, Imschloss, et al., 2024). By engaging the senses – sight, sound, touch, smell, and taste – service businesses are able to create an environment that enhances the overall customer experience (Hollebeek et al., 2024), thereby impacting consumer behavior (e.g. spending). Consider, for example, the strategic use of visual stimuli in a service – and retailscape design – as illustrated in the photographs of women's and



Figure 1. Retail environments with a focus on female clothing (left) and male clothing (right). Credits: Licht Kunst Licht AG.

men's clothing departments in two major retail stores in [Figure 1](#). While one environment features bright colors, emphasizing femininity to complement the female clothing offerings (left picture), the other environment employs a dark color scheme to highlight the masculinity aligned with the male clothing offerings (right picture).

Previous research has provided robust evidence of sensory stimuli's effects in service contexts across various sensory modalities. For example, [Zlateva et al., \(2024\)](#) recently showed that consumers' enjoyment of touch and sound enhances the service experience for those who had an Autonomous Sensory Meridian Response during the service delivery. Focusing on scent, [Ardelet et al. \(2022\)](#) documented that the presence of an ambient scent could increase consumers' comfort with frontline employee interactions in service settings. Similarly, a service environment's soundscape is well-known to impact not only consumers' evaluation of service-intensive retail environments, as well as in-store behavior in retail settings (e.g. [Grewal et al., 2003](#); [Milliman, 1982](#)), but also consumer satisfaction in tourist contexts (e.g. [Liu et al., 2018](#)).

LLMs might hold considerable promise for studying sensory stimuli's influence on consumer behavior, as they could capitalize on the link between language and human perception, thus offering a means to derive insights into consumers' sensory expectations or experiences solely based on the use of language ([Marjieh et al., 2024](#)). Illustrating this link, consumers often use sensory language to describe abstract concepts when making references to the five senses, such as when using the metaphor 'soft-hearted' to convey prosocial tendencies ([Cheon et al., 2019](#)). Such associations might also be present in LLM training data.

To date, only three studies have used LLMs to explore aspects of sensory experiences and perception. [Siedenburg and Saitis \(2023\)](#) prompted GPT-3 to describe the sound of eight instruments (e.g. flute, oboe, trump, violin) for each low-register, mid-register, and high-register note along a total of 20 verbally anchored semantic dimensions (e.g. the dimension 'pure, clear, clean', the dimension 'deep, thick, heavy'). Comparing the average GPT ratings with those of a human sample revealed significant correlations between twelve of the twenty semantic sound description dimensions. [Marjieh et al., \(2024\)](#) evaluated correlations between human and LLM-based similarity judgments of two stimuli in certain sensory modalities (e.g. pitch, loudness, colors, taste) by prompting different GPT version; for example, via hex codes (e.g. Color 1: #ff0000, Color 2: #A020F0), frequencies in hertz (e.g. Note 1: 587.3295358348151 Hz, Note 2: 987.7666025122483 Hz), and decibels (e.g. Tone 1: 72.6 dB, Tone 2: 74.1 dB). Their results indicate high correlations between the synthetic and human pairwise similarity ratings for pitch, loudness, and colors ($r = .84$ to $.92$ for the best GPT model), while lower ones were observed for taste ($r = .54$) and timbre ($r = .42$). Finally, [Motoki et al., \(2024\)](#) investigated whether GPT mimics consumers' tendencies to associate tastes (e.g. sweet) with specific colors and shapes (e.g. pink and round) across three languages. Their results show that GPT generally captures color-/shape-/taste correspondences, with some language-specific and version-specific variations.

While these studies provide valuable insights into LLMs' ability to decode certain sensory descriptions, the extent to which they might be capable of mirroring human responses to sensory stimuli – including the interpretation of sensory input to form concepts of objects or relationships ([Coren, 2003](#)) – remains unclear. LLMs' ability to, for example, decode hex codes of different colors and to assess their similarity, might not

correspond to a human's perception of colors. That is, they might not mimic human samples in the way they process, respond to, or are influenced by sensory stimuli, such as different colors. For example, would LLMs, in terms of color associations infer that a service provider has a more exciting (vs. competent) brand personality if the brand logo is red (vs. blue), as human samples showed (Labrecque & Milne, 2012)? Or, in terms of multisensory (in-)congruence, would LLMs mimic humans' tendency to assess servicescapes more favorably if the sensory stimuli (e.g. ambient scent and sound, or flooring and sound) are congruent rather than incongruent (Imschloss & Kuehnl, 2017; Mattila & Wirtz, 2001)?

As it is still unclear whether LLMs, like GPT-4o, could mimic the sensory stimuli's effects on consumer behavior reliably, our study sets out to shed light on these and similar questions. Specifically, we explore GPT-4o's ability to grasp color associations (Exploration 1), crossmodal correspondences (Exploration 2), and multisensory congruence effects (Exploration 3) that are well-established phenomena in sensory consumer research (Krishna, 2012; Labrecque & Milne, 2012; Spence, 2011) and service research (Mari & Poggesi, 2013). We do so by confronting GPT-4o with different sensory stimuli or scenarios, and testing whether its outputs align with color psychology's predictions (Wright, 1995), crossmodal correspondences (Spence, 2011), gestalt psychology (Koffka, 1935), and related fields. Aligning with prior research on LLMs (e.g. Y. Chen et al., 2023), we focus on long established effects in previous research with human participants, thereby dispensing with the need to compare each output with a new human sample.

Our insights add to research on sensory marketing and should therefore serve as a basis for discussion, promoting further explorations of GPT's capabilities to mimic or predict sensory stimuli's influence on consumer behavior in service settings. In this regard, our work offers various future research avenues that might stimulate further research on LLMs' ability to study aspects of sensory marketing in service settings. In terms of service practice, our aim is to inspire and guide managers in the service domain on when and how using GPT might be useful for pretesting sensory adjustments, or designing sensory elements of service environments.

2. GPT-Exploration 1: color associations

While LLMs might be capable of extracting and recoding sensory stimuli's structural properties (e.g. by analyzing color hues or saturation levels' of visual stimuli), it remains unclear whether they capture the associative meaning linked to sensory information, such as color hues. To shed light on this issue, we examined LLMs' tendency to mirror conceptual color associations in terms of the link between colors and the perceptions of a service provider's brand personality (Explorations 1a and 1b). Specifically, we focus on humans' associations of red with excitement and blue with competence, which have long been established in color psychology (e.g. Walters et al., 1982; Wexner, 1954; Wright, 1995), and have also been replicated in different marketing contexts (e.g. Labrecque & Milne, 2012). Such color associations are important for brand managers, as they serve as a basis for establishing visual brand identities (Jin et al., 2019). For example, companies use red logos to grab consumers' attention and to convey a feeling of excitement and energy (e.g. Coca-Cola, Netflix, Nintendo; Jin et al., 2019).

2.1. Exploration 1a: stand-alone assessments of color associations

We prompted GPT-4o with one of two differently colored logos in two separate trials. We asked GPT to identify the logo's color to test its multi-modal ability to identify colors correctly. We further asked it to rate the brand's competency on a scale from 1 ('not very competent') to 7 ('very competent'). We decided that this prompt structure would be similar to the questions and scales typically used in marketing research.¹

Figure 2 shows that GPT-4o correctly identifies the logos as either being red or blue; however, GPT-4o does not make an explicit reference to color associations. Instead, GPT-4o alludes to the logo's geometric shape and its overall impression, noting that the 'design looks somewhat minimalistic and abstract' (red logo), or that the 'design is simple and clear' (blue logo). In both instances, GPT-4o concludes that the overall design fails to convey 'a strong sense of professionalism or expertise'. Not mentioning – and perhaps even neglecting – the possible color associations with red or blue is surprising, as they are widely documented on marketing web pages and in blogs. In light of the convergent conclusion that the blue, as well as the red-colored logos, do not communicate professionalism or expertise well, it is somewhat surprising that GPT-4o's numerical rating of competence is slightly higher in respect of blue (rated 4) than red (rated 3). Although this rating points in the intended direction based on what one would expect from the respective color association documented in previous research (e.g. Walters et al., 1982; Wexner, 1954; Wright, 1995), the difference in ratings is minimal and might be a result of GPT's limited ability to return accurate numerical comparisons and evaluations (McCoy et al., 2023; Sarstedt, Adler, et al., 2024).

2.2. Exploration 1b: comparative assessment of color associations

While Study 1a demonstrated that GPT-4o only grasps the blue-competent association in a stand-alone assessment slightly, this result might become more salient in a comparative evaluation of two logos. That is, a direct comparison of the differently colored logos might bring color associations not evident in the stand-alone assessment to the fore.

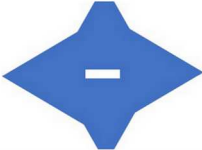
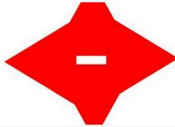
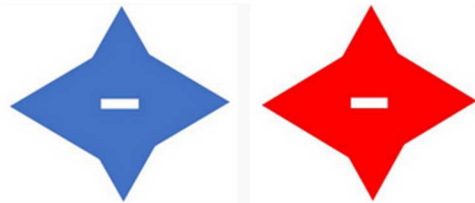
GPT-4o prompt and answer for the blue logo	GPT-4o prompt and answer for the red logo
<div style="text-align: center;">  </div> <p>Please imagine you are searching for a travel agency to book a summer vacation. You look at one travel agency on the internet. Attached you find the logo of this travel agency. Please rate on a scale from 1 to 7 how competent you think this travel agency is, with 1 = not very competent and 7 = very competent. Please state a numerical value for your rating and name the color of the logo.</p> <p>⊙ The color of the logo appears to be blue, with a white dash symbol in the middle. As for the rating of competence, based on the logo alone, I would rate it a 4 on a scale from 1 to 7. This is because the design is simple and clear, but it does not communicate a strong sense of professionalism or expertise. It is somewhat neutral in its impression, but the geometric shape could suggest modernity or a dynamic approach, which balances the simplicity.</p>	<div style="text-align: center;">  </div> <p>Please imagine you are searching for a travel agency to book a summer vacation. You look at one travel agency on the internet. Attached you find the logo of this travel agency. Please rate on a scale from 1 to 7 how competent you think this travel agency is, with 1 = not very competent and 7 = very competent. Please state a numerical value for your rating and name the color of the logo.</p> <p>⊙ Based on the image you uploaded, I see a red diamond-shaped logo with a white horizontal bar in the center. The logo uses bold, simple shapes and a striking red color.</p> <p>Here's my assessment:</p> <ul style="list-style-type: none"> • Color: The primary color of the logo is red. • Numerical Rating (Perceived Competence): The design looks somewhat minimalistic and abstract, which might not convey a strong sense of professionalism or expertise at first glance. It could give a modern or tech-oriented feel, but as a logo for a travel agency, it might lack a clear connection to travel or competence in the field. <p>Therefore, I would rate this logo 3 out of 7 in terms of perceived competence for a travel agency.</p>

Figure 2. Logo designs (separate prompts with blue logo (left) vs. red logo (right), stand-alone assessment).

We provided GPT-4o with differently colored logos next to one another, and asked the model to rate the logo that conveys a more competent brand on a 7-point scale. [Figure 3](#) shows that GPT-4o correctly identified the logo colors, and now makes an explicit reference to color associations. That is, GPT-4o alludes to blue's color associations ('Blue often conveys trust, stability, and professionalism, which could suggest reliability and competence (...)') and also offers an association related to red ('Red, while vibrant and attention-grabbing, is associated more with energy, urgency, or excitement but can sometimes evoke feelings of risk or caution'). It seems that when the logos are displayed next to each other, GPT-4o no longer focuses on the the logo's holistic impression (e.g. shape), but on the unique criteria that distinguish the two logos (e.g. the colors). Interestingly, GPT-4o justifies its rating of the blue-colored logo as being more competent by directly referencing color associations ('Considering this, I would rate TA1 as slightly more competent than TA2 based on these color associations').

Together, the results of Studies 1a and 1b suggest that while GPT-4o correctly identifies color hues, it does not explicitly use their associative meaning in stand-alone assessments.



Please imagine you are searching for a travel agency to book a summer vacation. You look at travel agencies on the internet. Attached you find the logo from two travel agencies (on the left is the logo of TA1 and on the right is the logo of TA2). Please rate on a scale from 1 to 7 which travel agency you perceive as more competent, with 1 = TA1 is more competent and 7 = TA2 is more competent. Please state a numerical value for your rating and name the colors of the two logos.



The logo on the left (TA1) is blue, and the logo on the right (TA2) is red.

Regarding the perceived competence, based solely on the logos:

- Blue often conveys trust, stability, and professionalism, which could suggest reliability and competence in a travel agency context.
- Red, while vibrant and attention-grabbing, is associated more with energy, urgency, or excitement but can sometimes evoke feelings of risk or caution.

Considering this, I would rate TA1 as slightly more competent than TA2 based on these color associations.

My rating on the scale would be 2, meaning TA1 is more competent but not overwhelmingly so.

Figure 3. Logo designs (single prompt with blue logo (left) and red logo (right), comparative assessment).

However, once provided with a comparative assessment of logos that only vary in color, GPT-4o relies explicitly on color associations for its judgment. Critically, despite alluding to extremely contrasting color associations, GPT-4o seems hesitant to transfer the magnitude of the contrast to its numerical judgments. In addition, GPT-4o does not clearly discriminate the concepts mentioned in the color associations, which might prove problematic in empirical follow-up assessments (Voss et al., 2020).

3. GPT-Exploration 2: crossmodal correspondences

Human senses do not function in isolation, but are inherently intertwined in the form of crossmodal correspondences, which refer to systematic associations between information across the senses, so that inputs in one sensory modality might influence perceptions or responses in another sensory modality (Chuquichambi et al., 2024; Spence, 2011; Spence et al., 2009; Velasco et al., 2015). For example, crossmodal correspondences might occur between consumers' auditory and haptic perceptions (Imschloss & Kuehnl, 2019), auditory and taste perceptions (Crisinel et al., 2012; see Guedes et al., 2023 for a review), as well as color and taste perceptions (Spence & Levitan, 2021; Sugimori & Kawasaki, 2022). As an example, Imschloss and Kuehnl (2019) show that consumers perceive fabrics to be softer when listening to soft (vs. hard) music in retail environments.

A prominent stream of research in this field examines shape and sound symbolism effects, showing that particular sounds, as well as specific abstract shapes, are associated with certain other sensory qualities (Spence, 2012). The kiki-bouba effect in sound symbolism, which is rooted in gestalt psychology (Koffka, 1935), is a particularly prominent example of such crossmodal shape-sound associations. According to this effect, humans tend to match the speech sound of the fictive word 'kiki' with spiky angular shapes and the speech sound of 'bouba' with round shapes (Y.-C. Chen et al., 2016; Ćwiek et al., 2021). Previous research substantiated that the kiki-bouba effect is a fundamental crossmodal correspondence in human perception that develops in early childhood (Fort et al., 2018) and generalizes across various languages and cultures (Ćwiek et al., 2021).

The kiki-bouba effect has important implications for servicescape design, as it highlights the interaction between two sensory stimuli (Mari & Poggesi, 2013). For example, the shapes used in physical settings, such as architecture, interior design, and signage, should align with the sound of the business's name. A soft sounding restaurant name would feel mismatched if the interior is filled with sharp, angular furniture. Similarly, for service industries focused on relaxation (like spas), using 'bouba'-inspired design elements (soft names, round shapes) could help create a calming atmosphere, thereby aligning the auditory and the visual elements to enhance the sense of relaxation.

To examine whether LLMs also demonstrate an association between the words kiki/bouba and abstract angular vs. round shapes, we prompted GPT-4o with a spiky shape and a round shape (Figures 4 and 5). Instead of presenting the words as auditory stimuli, we presented them visually as part of the prompt itself. That is, for each of the abstract shapes, we asked whether the shape would be associated with the name 'kiki' (1 on a 7-point scale) or with 'bouba' (7 on a 7-point scale). We again asked GPT-4o for separate and comparative evaluations.










GPT-4o prompt and answer for the angular/kiki shape	GPT-4o prompt and answer for the round/bouba shape
<div style="text-align: center;">  </div> <p data-bbox="322 365 625 412">Please label the attached shape. To label the shape please indicate on a scale from 1 to 7 which name you would give it, with 1 = it's name is Kiki to 7 = it's name is Bouba. Please provide a numerical value as an answer.</p> <p data-bbox="184 440 611 515">  Based on the visual characteristics of the shape you provided, it has sharp angles and an irregular, jagged form. Shapes with sharp angles are typically associated with the name "Kiki" in psychological studies regarding sound symbolism, where "Kiki" is linked to spiky, angular shapes, and "Bouba" is associated with rounded, smooth shapes. </p> <p data-bbox="217 521 546 540">On a scale from 1 to 7, where 1 = Kiki and 7 = Bouba, I would label this shape as 1.</p> <p data-bbox="217 544 309 562">  </p>	<div style="text-align: center;">  </div> <p data-bbox="796 356 1085 403">Please label the attached shape. To label the shape please indicate on a scale from 1 to 7 which name you would give it, with 1 = it's name is Kiki to 7 = it's name is Bouba. Please provide a numerical value as an answer.</p> <p data-bbox="664 431 1085 478">  Based on the provided shape, which appears rounded and smooth with no sharp angles, it aligns more with the characteristics typically associated with the name "Bouba" in the Bouba-Kiki effect, a psychological phenomenon linking sounds to shapes. </p> <p data-bbox="690 487 1059 515">On the scale from 1 (Kiki) to 7 (Bouba), I would label this shape as a 6. The smooth and rounded contours evoke the softer sounds of "Bouba" rather than the sharp sounds of "Kiki".</p> <p data-bbox="690 525 783 544">  </p>

Figure 4. Kiki-bouba effect (separate prompts, stand-alone assessment).

Please label the two attached shapes. To label the shape please indicate on a scale from 1 to 7 which name you would give to the left as well as right shape, with 1 = it's name is Kiki to 7 = it's name is Bouba. For each shape (left and right), please provide a numerical value as an answer.

 Based on the shapes you provided, this seems to reference the well-known "Bouba-Kiki Effect" study, where people tend to associate sharp, jagged shapes with the name "Kiki" and rounded, blobby shapes with the name "Bouba."

- **Left shape** (jagged, sharp): People typically identify this as **Kiki**. On a scale from 1 (Kiki) to 7 (Bouba), I'd rate this as 1.
- **Right shape** (rounded, blobby): People commonly associate this with **Bouba**. On the same scale, I'd rate this as 7.

So, the numerical values would be:

- Left shape: 1
- Right shape: 7

Let me know if you need further clarification!




Figure 5. Kiki-bouba effect (single prompt, comparative assessment).

Figures 4 and 5 show that GPT-4o indeed mimics correspondences regarding both the spiky 'kiki' shape and the round 'bouba' shape in both the stand-alone as well as the comparative assessment. In this exploration, GPT-4o also exhibits meta-knowledge by referring directly to the kiki-bouba effect in psychological literature.

4. GPT-Exploration 3: multisensory congruence

Humans experience servicescapes holistically, incorporating information from multiple senses simultaneously (Mari & Poggese, 2013). Prior research has shown that multisensory congruence – that is, stimuli that match (vs. mismatch) in terms of their arousal level, in their semantic-perceptual level, or in their thematic meaning – lead to positive consumer reactions (e.g. Imschloss & Kuehnl, 2017; Spence et al., 2014). For example, matching (vs. mismatching) the arousal level of in-store music and scent increased consumers' satisfaction with the shopping experience, and led to more favorable in-store behaviors like impulse purchases and approach behaviors (Mattila & Wirtz, 2001). Similarly, matching – as opposed to mismatching – haptic stimuli such as flooring and background music along the semantic-perceptual dimension of softness has been shown to have positive effects on product evaluations in retail environments (Imschloss & Kuehnl, 2017). Multisensory congruence can also be established along thematic dimensions, as evidenced by research documenting more favorable attitudes toward a mock retail environment employing combinations of Christmas scent and Christmas music when compared to combinations of no scent and Christmas music (Spangenberg et al., 2005). Similarly, in respect of food consumption environments, previous research has shown a higher than expected liking of the food when the food products are consumed in a thematically congruent consumption context, such as consuming a watermelon (vs. chocolate truffle) in a congruent virtual summer (vs. winter) eating environment, or consuming a popsicle (vs. sushi) in a congruent beach (vs. sushi restaurant) context (Schouteten et al., 2024; Van Bergen et al., 2021).

To test whether LLMs could help examine service or retail environments' effects, we used the servicescape settings reported in the study by Mattila and Wirtz (2001) as a template to create sensory scenarios. These authors conducted a 3 × 3 between-participant field study in a gift shop to examine the effects of multisensory congruence along the arousal dimension. They did so by either dispersing grapefruit ambient scent (high arousal), or lavender ambient scent (low arousal), or no scent but employed classical music by Elizabeth Miles by playing either the energizing collection featuring fast-tempo music (high arousal), or the relaxing collection encompassing slow-tempo music (low arousal) or no music. Their results showed that the matching (vs. mismatching) arousal qualities of the two ambient sensory stimuli guide consumers to exhibit higher levels of approach behavior, impulse buying, and satisfaction with the shopping experience.

Building on this study design, we prompted GPT-4o with four sensory scenarios that describe servicescapes as being scented with a low or high arousal scent (i.e. lavender, grapefruit) and employing low or high arousal music (i.e. slow vs. fast tempo classical music). We asked GPT-4o to indicate a consumer's satisfaction with the experience in the respective servicescape, since this variable's interaction effect had the highest effect size in the original study compared to the other outcome variables.

GPT-4o did not provide any indication of arousal (in)congruence effects on satisfaction (Table 1). To illustrate, GPT-4o rated consumers' satisfaction level equally as a five on a seven-point scale when prompting the arousal incongruent servicescape scenario with fast music and lavender scent, as well as when prompting the arousal congruent servicescape scenario with fast music and grapefruit scent. Likewise, there was no difference in GPT-4o's ratings of the arousal-incongruent servicescape with slow music and lavender

Table 1. Prompt and GPT-4o reactions to multisensory congruence task.

GPT-4o prompt: Please envision the following servicescape: The servicescape employs *fast/slow* tempo classical music and a *lavender/grapefruit* ambient scent. Please indicate a customer's satisfaction with the experience in this servicescape on a scale from 1 to 7, with 1 = not very satisfied to 7 = very satisfied. Please provide a numerical value as an answer.

GPT-4o answers for different retail environment scenarios

	Lavender	Grapefruit
Fast tempo music	Given the description of the servicescape with fast tempo classical music and a lavender ambient scent, the customer's satisfaction can vary based on personal preferences. However, both classical music and lavender scent are commonly associated with positive effects on mood and relaxation, although the fast tempo of the music might counteract some of the calming effect of lavender. On a scale from 1 to 7, a reasonable assumption would place satisfaction around 5. This represents generally positive, but not necessarily overwhelming, satisfaction.	Based on the description of the servicescape, the combination of fast tempo classical music and a grapefruit ambient scent might create a stimulating and energetic environment. While individual preferences vary, a general assumption can be made that this type of atmosphere would appeal to those who enjoy lively environments and fresh, citrusy scents. Given these elements, I would estimate the customer's satisfaction with the experience in this servicescape as 5 out of 7. This reflects moderate to high satisfaction, assuming the customer has a positive response to both the music and scent combination. However, satisfaction could vary depending on personal tastes.
Slow tempo music	Given the described servicescape – slow tempo classical music and a lavender ambient scent, both of which are commonly associated with relaxation and positive mood enhancement – it's reasonable to infer that a customer would likely have a favorable response. These elements are often used in environments designed to promote calmness, comfort, and a pleasant atmosphere. On a scale from 1 to 7, I would estimate a customer's satisfaction to be approximately 6. This reflects a high level of satisfaction, though individual preferences and context could shift this score slightly.	Based on the description of the servicescape, I would estimate a customer's satisfaction with the experience to be a 6 on a scale from 1 to 7. Slow tempo classical music and grapefruit ambient scent tend to evoke a calm, pleasant, and refreshing atmosphere, which is likely to enhance customer satisfaction, though individual preferences could lead to variations.
No music	Given the description of a servicescape with no music and a lavender ambient scent, customer satisfaction may vary depending on individual preferences. Since lavender is commonly associated with relaxation and calmness, the absence of music may further enhance a tranquil atmosphere. However, some customers may prefer background music to complement the ambiance. Considering these factors, I would estimate a moderate to high satisfaction rating. A reasonable assumption for customer satisfaction could be: 5 out of 7. This rating accounts for the likely positive influence of the lavender scent but also considers that the lack of music might not appeal to everyone.	The numerical value for a customer's satisfaction with the described servicescape would likely be around 4. This is based on the fact that while grapefruit is generally considered a pleasant and refreshing scent, the absence of music may make the environment feel neutral or lacking in ambiance for some customers. Satisfaction can vary greatly depending on individual preferences, but a neutral response of 4 seems reasonable for an experience that might feel incomplete without auditory elements.

Note: The words in italics are provided depending on the condition.

scent, and the servicescape with slow music and grapefruit scent. Here again, GPT expects consumers to have an equal satisfaction level (six on a seven-point scale). We also find that the descriptions of service environments with two sensory stimuli (music and scent) did not produce substantially different results than descriptions only appealing to one sensory stimulus (scent). Specifically, GPT-4o returns a generic response with a favorable assessment of olfactory stimuli per se, and does not refer to distinctive interactive effects that might be stimulated by slower vs. faster music, and less vs. more arousing scents. Instead, describing servicescapes that employ either a grapefruit or lavender scent without music, leads GPT-4o to generically suggest that consumers in such servicescapes could have an 'experience that might feel incomplete without auditory elements'. Overall, the output indicates that the congruence between the arousal levels of different music tempi and scent types is not readily available to the model; possibly because such associations are often not described explicitly, could therefore not be inferred from the training data.

Apart from the aforementioned stand-alone assessment, we next prompted GPT-4o with comparative servicescape scenarios. Specifically, we prompted GPT-4o with both an incongruent and a congruent servicescape scenario, and asked it to indicate in which servicescape consumers might be more satisfied with their experience. Unlike in the stand-alone assessment, GPT-4o now reveals indications of multisensory congruence effects. Rendering this understanding of multisensory (in)congruence, [Figures 6 and 7](#) document that for servicescape scenarios of slow music paired with (incongruent) grapefruit versus (congruent) lavender scent, as well as for servicescape scenarios of fast music paired with (incongruent) lavender versus (congruent) grapefruit scent, GPT-4o suggests higher satisfaction ratings with the respective congruent servicescape conditions. Specifically, GPT-4o explicates that 'a consumer is likely to be more satisfied with Servicescape 1 due to the calming synergy of lavender and slow tempo classical music' ([Figure 6](#)). For the fast music scenarios, GPT-4o notes that 'a consumer might be more satisfied in Servicescape 2, as the grapefruit scent complements the fast tempo classical music by enhancing feelings of energy and alertness' ([Figure 7](#)). Overall, GPT-4o accurately describes the environment with the potential associations that come from both the auditory and the olfactory stimuli. The model also correctly returns the (in)congruence effects on a descriptive level, but once again does not translate them into substantially differing numerical evaluations.

5. Discussion

5.1. Theoretical implications and future research directions

This article contributes to the ongoing discussion on AI's applicability in market research, specifically setting the stage for future explorations into the capabilities and limits of LLMs regarding mimicking human responses to sensory stimuli and their influence on consumer behavior in service settings. Across three explorations, we find that LLM-generated synthetic datasets represent a promising frontier in service research, especially given the most recent development of multimodal AI interfaces like GPT-4o.

Our exploration suggests that LLMs, like GPT, possess the potential to mimic certain human responses to sensory stimuli, especially those reflected in language. GPT was

Please envision the following two servicescapes. Servicescape 1: The servicescape employs slow tempo classical music and a lavender ambient scent. Servicescape 2: The servicescape employs slow tempo classical music and a grapefruit ambient scent. Please indicate a customer's satisfaction with the experience in these two servicescapes on a scale from 1 to 7, with 1 = not very satisfied to 7 = very satisfied. Please provide a numerical value as an answer and state in which servicescape a consumer might be more satisfied with the experience.



Based on research into sensory marketing and the impact of music and scent on customer satisfaction, we can make an educated assumption regarding the potential satisfaction in these two servicescapes:

1. **Servicescape 1 (slow tempo classical music + lavender scent):** Lavender is known for its calming and relaxing effects. Combined with slow tempo classical music, this creates an environment conducive to relaxation and comfort, which is generally favorable in service settings like spas, lounges, or upscale retail environments. I would estimate a satisfaction score of 6 on a scale of 1 to 7 for this servicescape, as it promotes a calming, positive experience.
2. **Servicescape 2 (slow tempo classical music + grapefruit scent):** Grapefruit scent, while fresh and invigorating, is often associated with alertness and energy. While it may still create a pleasant atmosphere, it could be slightly less consistent with the relaxing effect of slow tempo classical music. As a result, a consumer might find the combination somewhat less harmonious than the lavender pairing. I would estimate a satisfaction score of 5 for this servicescape, as it might still be enjoyable but less cohesive.

Conclusion: A consumer is likely to be more satisfied with **Servicescape 1** due to the calming synergy of lavender and slow tempo classical music.



Figure 6. Servicescape scenario with slow music and a lavender scent (congruent sensory stimuli) versus a servicescape scenario with slow music and a grapefruit scent (incongruent sensory stimuli).

especially helpful for the direct comparisons of sensory stimuli, for example, those concerning color associations (Exploration 1a and 1b) and scenarios of multisensory congruence (Exploration 3). For sound symbolism (Exploration 2), GPT-4o performed satisfactorily in separate as well as comparative assessments. These findings relate to recent advances in sensory research with LLMs (Motoki et al., 2024) and in the field of silicon sampling (Arora et al., 2025; Sarstedt, Adler, et al., 2024). Specifically, we extend the work of Motoki et al. (2024), who found that GPT associates certain colors and shapes with specific tastes by demonstrating that GPT-4o also associates shapes with sounds – a crossmodal correspondence long established in studies with human participants.

However, our results also show that GPT-4o's ability to portray sensory experiences diverges from humans in some instances – at least in its current implementation. GPT-4o, for example, did not identify color associations and multisensory congruence effects when the different stimuli were presented separately in stand-alone assessments.

Please envision the following two servicescapes. Servicescape 1: The servicescape employs fast tempo classical music and a lavender ambient scent. Servicescape 2: The servicescape employs fast tempo classical music and a grapefruit ambient scent. Please indicate a customer's satisfaction with the experience in these two servicescapes on a scale from 1 to 7, with 1 = not very satisfied to 7 = very satisfied. Please provide a numerical value as an answer and state in which servicescape a consumer might be more satisfied with the experience.



For these two servicescapes, customer satisfaction could be influenced by how the ambient scent and music impact mood and sensory experience.

Servicescape 1: Fast tempo classical music and a lavender ambient scent.

- **Lavender** is commonly associated with relaxation and stress relief. When combined with fast tempo classical music, this creates a somewhat contrasting sensory experience, as the music may energize, while the scent calms.
- Estimated satisfaction score: 5 (Some customers may find this combination unusual or conflicting, but others may appreciate the balance between energy and relaxation).

Servicescape 2: Fast tempo classical music and a grapefruit ambient scent.

- **Grapefruit** is often linked to feelings of freshness and alertness. When paired with fast tempo classical music, this combination is more likely to energize and uplift, creating a coherent and stimulating environment.
- Estimated satisfaction score: 6 (This combination is more likely to align with the invigorating effect of the music, leading to higher satisfaction).

Conclusion:

A consumer might be more satisfied in **Servicescape 2**, as the grapefruit scent complements the fast tempo classical music by enhancing feelings of energy and alertness.

Figure 7. Servicescape scenario with fast music and a lavender scent (incongruent sensory stimuli) versus a servicescape scenario with fast music and a grapefruit scent (congruent sensory stimuli).

The observed divergencies give rise to three future research areas that might address questions pertaining to (i) LLMs' responses to sensory information in general, (ii) LLMs' representation of crossmodal correspondences and multisensory congruence, and (iii) ways to use LLMs in sensory consumer studies.

5.1.1. How do LLMs respond to sensory information?

Further research is required to examine the conditions under which LLMs are capable of describing sensory experiences and more complex multisensory interplays that characterize human perception of servicescapes (Spence et al., 2014). Extending our example on the role of color associations and brand personality ratings, future research should

systematically examine the potential of LLMs to mimic a diverse set of human sensory associations as well as their influence on consumer behavior. For example, future research could examine whether LLMs adequately represent the associative meaning of sensory descriptions regarding various colors, music, and scents. Do LLMs, for example, consider a jasmine scent as rather relaxing? And if so, are LLMs capable of mimicking the nonconscious positive effect of jasmine scents on service quality evaluations, as Girard et al. (2019) have shown in the context of public transportation services? Do LLMs associate a coffee scent with alertness and with improvements in (expected) performance levels as shown in respect of human samples (Madzharov et al., 2018)? Or – in a service context – do LLMs discern national associations in music (e.g. Spanish vs. Italian music) and mimic the influence of national-specific music on consumer behavior in a restaurant setting (e.g. preferences for seafood paella vs. chicken parmesan), as documented in previous research (Zellner et al., 2017)?

In a related context, it would be interesting to explore whether LLM responses to sensory information are sensitive to the specific servicescape types that are frequently discussed in the literature (Mari & Poggesi, 2013), such as banks, doctors' waiting rooms, restaurants, and sports venues. For example, Kabadayi et al. (2023, p. 516) emphasize the significance of sensory and environmental factors in creating brand love in restaurant settings, thereby highlighting the particular role of servicescape characteristics such as 'attractive wall decorations, pleasing plants/flowers, quality furniture, colors that create a warm atmosphere, relaxing and enjoyable music, soothing fragrance, well-adjusted room temperature, warm lighting, tasteful and high-quality table setting, spacious dining areas, and a sufficient number of well-groomed service staff'. Future research could thus address a wide range of questions specific to particular service settings, such as whether LLMs would differentiate between service settings in which a calming atmosphere would be preferable (e.g. spas or doctors waiting rooms) versus settings in which consumers would look for stimulation (e.g. restaurants or sports venues). Other research might examine whether LLMs could mirror human responses to specific sensations, like thermo-perception, accurately, or whether LLMs could effectively capture and represent responses to sensory elements that are particularly meaningful in certain servicescapes when provided with pictorial or written information, such as the interior design of restaurants. Likewise, future research could prompt GPT with pictures or videos of differently crowded service and retail settings to assess their suitability as mirrors of human responses to crowding or to other customers as documented in previous service research (e.g. Ittamalla & Srinivas Kumar, 2019; Li et al., 2009; Uhrich & Benkenstein, 2012). This approach might also facilitate the examination of diverse congruence effects between music played in service environments and other servicescape elements such as consumer density or the generic service category (Oakes & North, 2008).

5.1.2. How do LLMs represent crossmodal correspondences and multisensory congruence?

In light of the robust empirical evidence of crossmodal correspondences and the effects of multisensory (in)congruence (Sarstedt, Im Schloss, et al., 2024), future research should explore LLMs' capability to adequately represent the human-like connections between different sensory modalities.

First, in the domain of crossmodal correspondences, future research could investigate for which sensory modalities LLMs produce human-like correspondences. Color-taste correspondences – as is common in restaurant settings – have recently been investigated (Motoki et al., 2024), although researchers need to acknowledge that not all crossmodal correspondences could be readily examined with LLMs. To illustrate, crossmodal effects between auditory stimuli and taste perceptions showing that, for example, bitter, sweet, or zingy and refreshing-sounding music modulate taste perceptions (Crisinel et al., 2012; North, 2012), can only scantily be verbalized in the text with which the model was trained. Cross-modal associations are likely to only occur in LLMs if basic knowledge of individual associations was represented in their training data. GPT-explorations 2 and 3 specifically show that GPT-4o refers directly to the kiki-bouba-effect as well as to the already established calming properties of the lavender scent and the grapefruit scent's arousing properties. GPT might therefore be less capable of mirroring human responses to sensory stimuli or to atmospheric compositions less represented in the training data.

To systematically explore LLMs' ability to mirror crossmodal correspondences, future research could draw on Spence's (2011) categorization of crossmodal correspondences. While some crossmodal correspondences might occur due to the correlations of stimulus dimensions in the natural environment (e.g. the correlation between object size and resonance frequency), others might occur due to shared neuro-structural representations in the brain (e.g. for magnitude-related dimensions like loudness and size) or due to a semantic overlap in stimulus description along different sensory dimensions (e.g. the words 'low' and 'high' can be used to describe visual elevation or auditory pitch) (Spence, 2011). It would be interesting for future research to examine whether LLMs are particularly effective in mimicking semantically mediated crossmodal correspondences due to their grounding in language as opposed to statistical or structural crossmodal correspondences.

Second, given the effects of multisensory (in)congruence on consumer decision-making (Sarstedt, Imschloss, et al., 2024), it would also be interesting to examine whether LLMs identify when and which sensory stimuli are (in)congruent. Specifically, are LLMs equally capable of identifying different types of congruence, such as multisensory congruence along the arousal, semantic-perceptual, and thematic dimensions? Would GPT, for example, be useful to mirror human responses in terms of expected food preferences in different food-context combinations (van Bergen et al., 2021), such as when GPT is prompted with thematically congruent food-context settings (popsicle at the beach, sushi in the sushi restaurant) as opposed to incongruent settings (popsicle in the sushi restaurant, sushi at the beach)? It would be plausible to assume that LLMs are particularly effective in capturing multisensory (in)congruence for stimuli that (mis)match along the semantic-perceptual or thematic dimension. Future research could therefore explore whether LLMs are indeed less capable of mimicking multisensory (in)congruency effects of stimuli that (mis)match in their psychological or physiological arousal level, as these dimensions might be less represented in the underlying text corpus.

5.1.3. How to use LLMs for sensory consumer studies?

Across all domains discussed above, studying the effectiveness of LLMs in mimicking the influence of sensory cues on consumer behavior in service contexts, requires an

understanding of how they need to be instructed to produce human-like responses – for example, in terms of prompting techniques that induce a certain level of the results' variability. Our findings suggest that the prompting technique (i.e. using comparative rather than stand-alone assessments) might help generate accurate sensory descriptions. Researchers could therefore systematically examine the implications of testing sensory stimuli in isolation versus in comparative settings (e.g. testing the impact of different sensory food descriptions in isolation vs. next to each other; Garaus et al., 2023). Also, future research should validate prompting techniques that account for previous human sensory experiences. To illustrate, a consumer's previous experience with certain foods or beverages in restaurant settings, and in keeping with this consumer's evaluation of food or beverage pairings might depend strongly on sensory familiarity as well as on previous sensory experiences with the combined foods or beverages (Van Bergen et al., 2022). Future research might also explore how to best infuse LLMs with some types of previous human sensory experiences, and examine whether, for example, a food pairing of soy sauce and vanilla ice cream is preferred when prompted with a description of a consumer who has consumed soy sauce multiple times versus never before. Taking this example further: would it be plausible to assume or to test the effects of repeated exposure or sensory habituation by prompting LLMs accordingly? In addition, given that most sensory experiences can only be examined by using language descriptions of sensory impressions, more research is needed on how to best prompt such 'sensory scenarios' of servicescape experiences. For example, future research should assess if LLMs' ability to produce human-like responses to sensory stimuli depends on whether the stimuli are operationalized in terms of perceptual effects (e.g. 'loud music' or 'intense color') as opposed to their physical parameters (e.g. '80 dB' or '100% saturation').

In terms of generating human-like output, future research should further examine whether LLMs differ in their capability to mimic different sensory effects' dimensions. Since sensory cues could influence consumers' affective (e.g. mood), physiological (e.g. arousal), cognitive (e.g. thoughts), and behavioral (e.g. approach behavior) responses, the question arises which of these experiential dimensions are adequately reflected in LLM-generated synthetic data. Specifically, researchers could explore which particular outcome measures LLMs mimic well. While it might, for instance, be more difficult to use LLMs to predict sensory stimuli's influence on outcomes, such as approach behavior and impulse purchases, they might be better equipped to approximate human responses, such as satisfaction, willingness-to-pay, and purchase intentions, which are probably more present in the underlying text corpus. Finally, using LLMs as a catalyst for extracting sensory descriptions and associations from the training data might provide a means to generate hypotheses about crossmodal correspondences or congruence effects. Such associations could then be tested in subsequent studies with a human validation sample (Sarstedt, Adler, et al., 2024).

5.2. Practical implications

Regarding practical implications, our explorations underscore the potential of GPT-4o as an initial tool for managers designing the sensory aspects of service environments. For instance, GPT-4o captures fundamental effects of sound symbolism (Exploration 2), aiding managers in selecting names and shapes (e.g. for furniture) that align with

desired sensory impressions. Additionally, our findings demonstrate that GPT-4o is particularly valuable for comparative assessments, such as identifying specific color associations (Exploration 1b) and evaluating the multisensory (in-)congruence of ambient, stimuli-like scent and music (Exploration 3).

In other words, GPT-4o could help establish which sensory properties harmonize with other stimuli or semantic content when comparing different options. This makes it a useful pretesting tool for decisions related to brand positioning, color schemes, and the interplay of multiple sensory elements in service environments. By being both fast and cost-effective, GPT-4o enables managers to conceptually evaluate and refine sensory designs. Since much sensory information is likely embedded in LLMs' training data, GPT-4o could transform this knowledge into actionable text, thereby offering a practical way of initially assessing sensory elements in service contexts.

Overall, our explorations suggest that GPT-4o could act as a versatile tool for managers in the service industry, helping them assess the effects of sensory stimuli, and anticipate consumer responses to sensory elements or environments in service contexts. LLMs, like GPT-4o, enable the fast and cost-effective pretesting of various sensory stimuli, and their potential impacts.

However, our findings also indicate that GPT-4o occasionally struggles to accurately mimic color associations and assess multisensory congruence when sensory stimuli are described individually rather than compared directly (i.e. when GPT-4o is asked to evaluate two or more specific sensory descriptions side by side). Managers intending to use GPT-4o for pretesting sensory elements in their service environments are therefore advised to avoid stand-alone evaluations. Instead, they should create comparative descriptions of alternative environments for GPT-4o to analyze and evaluate more effectively.

5.3. Limitations

Our manuscript describes three explorations into GPT's ability to mimic human responses to sensory stimuli, serving as a starting point for more systematic assessments of sensory silicon sampling. However, our approach comes with several limitations, including a focus on a single model (GPT-4o) rather than a comprehensive evaluation of responses from multiple available LLMs (e.g. Llama, Claude, Gemini, etc.). Additionally, we did not explore variations in model settings, such as temperature, top-k, or top-p.

Although it is multimodal, GPT-4o can only process language and visual stimuli, but not olfactory, auditory, gustatory, or haptic stimuli. The inherent boundaries of the interface's perceptual capabilities thus limit LLMs' ability to assess crossmodal correspondences' and multisensory congruence's effects – at least for the time being. Modeling phenomena of crossmodal correspondences or multisensory congruence thus is currently only possible through written descriptions of sensory scenarios or the use of pictorial material. While this naturally restricts potential research into the areas of sensory expectations and associations, our findings nonetheless offer insights into the extent to which LLMs respond to sensory input just like human samples do.

We further decided to examine well-known effects in sensory research as documented by previous research based on human participants, thereby dispensing with the need to contrast each output with a specific human sample. While the intention was to examine if

GPT-4o could mimic basic sensory effects, descriptions of the findings were probably included in the training data. This might not harm GPT's ability to sufficiently reflect human assessments of known stimuli, but it might limit its usefulness regarding new and complex stimuli.

6. Conclusion

While many existing studies on AI implementation in service contexts have examined its ethical and responsible implications for employees and customers (Belanche et al., 2024; Tlili et al., 2024; Vatankhah et al., 2024; Weng et al., 2024), this paper contributes to our understanding of the opportunities and limitations of AI's application in market research. Specifically, we regard our article as a starting point for future research to systematically explore the effectiveness of LLMs regarding mimicking human-like responses to sensory stimuli as commonly encountered in service settings. Such exploration should focus on examining how LLMs process sensory input and deal with the peculiarities of human sensory information processing, such as crossmodal correspondence and multisensory (in)congruence. These explorations also require researchers to address the role of prompting strategies in generating and mimicking human-like responses to sensory stimuli.

Note

1. To facilitate reproducibility, we document all prompts in the figures (Deer et al., 2025).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Ethical statement

This research did not involve human participants. This study exclusively employed GPT-4o for data generation.

ORCID

Monika Imschloss  <http://orcid.org/0000-0001-8582-5965>

Marko Sarstedt  <http://orcid.org/0000-0002-5424-4268>

Susanne J. Adler  <http://orcid.org/0000-0002-3211-6871>

Jun Hwa Cheah  <http://orcid.org/0000-0001-8440-9564>

References

- Ahn, J., & Back, K.-J. (2020). The structural effects of affective and cognitive elaboration in formation of customer–brand relationship. *The Service Industries Journal*, 40(3–4), 226–242. <https://doi.org/10.1080/02642069.2018.1460358>
- Ardelet, C., De Peyrelongue, B., & Mérigot, P. (2022). Within sniffing distance: Impact of ambient scent and physical distancing on consumer comfort with frontline employees. *Journal of Personal Selling & Sales Management*, 42(3), 265–278. <https://doi.org/10.1080/08853134.2022.2049602>
- Arora, N., Chakraborty, I., & Nishimura, Y. (2025). AI-Human hybrids for marketing research: Leveraging LLMs as collaborators. *Journal of Marketing*, 89(2), 43–70. <https://doi.org/10.1177/00222429241276529>

- Belanche, D., Belk, R. W., Casalo, L. V., & Flavián, C. (2024). The dark side of artificial intelligence in services. *The Service Industries Journal*, 44(3–4), 149–172. <https://doi.org/10.1080/02642069.2024.2305451>
- Binz, M., & Schulz, E. (2023). Using cognitive psychology to understand GPT-3. *Proceedings of the National Academy of Science*, 120(6), Article e2218523120. <https://doi.org/10.1073/pnas.2218523120>
- Brand, J., Israeli, A., & Ngwe, D. (2023). *Using GPT for market research*. (Harvard Business School Marketing Unit Working Paper No. 23-062). SSRN. https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=4395751
- Chen, Y.-C., Huang, P.-C., Woods, A., & Spence, C. (2016). When “Bouba” equals “Kiki”: Cultural commonalities and cultural differences in sound-shape correspondences. *Scientific Reports*, 6(1), 26681. <https://doi.org/10.1038/srep26681>
- Chen, Y., Kirshner, S., Ovchinnikov, A., Andiappen, M., & Jenkin, T. (2023). *A manager and an AI walk into a bar: Does ChatGPT make biased decisions like we do?* SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4380365
- Cheon, B. K., Tan, K. M., & Lee, L. L. (2019). Tender food, tender hearts: The metaphorical mapping of hard-soft orosensory signals to interpersonal trust and prosocial tendencies. *Food Quality and Preference*, 71, 242–249.
- Chuquichambi, E. G., Munar, E., Spence, C., & Velasco, C. (2024). Individual differences in sensitivity to taste-shape crossmodal correspondences. *Food Quality and Preference*, 115, 105110. <https://doi.org/10.1016/j.foodqual.2024.105110>
- Coren, S. (2003). Sensation and perception. In D. K. Freedheim (Ed.), *Handbook of psychology: History of psychology* (pp. 85–108). John Wiley & Sons.
- Crisinel, A.-S., Cosser, S., King, S., Jones, R., Petrie, J., & Spence, C. (2012). A bittersweet symphony: Systematically modulating the taste of food by changing the sonic properties of the soundtrack playing in the background. *Food Quality and Preference*, 24(1), 201–204. <https://doi.org/10.1016/j.foodqual.2011.08.009>
- Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., Kawahara, S., Koutalidis, S., Krifka, M., Lippus, P., Lupyan, G., Oh, G. E., Paul, J., Petrone, C., Ridouane, R., Reiter, S., Schümchen, N., Szalontai, Á, Únal-Logacev, Ö, ... Winter, B. (2021). The bouba/kiki effect is robust across cultures and writing systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1841), Article 20200390. <https://doi.org/10.1098/rstb.2020.0390>
- Deer, L., Adler, S. J., Datta, H., Mizik, N., & Sarstedt, M. (2025). Toward open science in marketing research. *International Journal of Research in Marketing*, 42(1), 212–233. <https://doi.org/10.1016/j.ijresmar.2024.12.005>
- Demszky, D., Yang, D., Yeager, D. S., Bryan, C. J., Clapper, M., Chandhok, S., Eichstaedt, J. C., Hecht, C., Jamieson, J., Johnson, M., Jones, M., Krettek-Cobb, D., Lai, L., JonesMitchell, N., Ong, D. C., Dweck, C. S., Gross, J. J., & Pennebaker, J. W. (2023). Using large language models in psychology. *Nature Reviews Psychology*, 2, 688–701. <https://doi.org/10.1038/s44159-023-00241-5>
- Fort, M., Lammertink, I., Peperkamp, S., Guevara-Rukoz, A., Fikkert, P., & Tsuji, S. (2018). Symbolouki: A meta-analysis on the emergence of sound symbolism in early language acquisition. *Developmental Science*, 21(5), e12659. <https://doi.org/10.1111/desc.12659>
- Gao, Y., Lee, D., Burtch, G., & Fazelpour, S. (2024). *Take caution in using LLMs as human surrogates: Scylla ex machina*. arXiv:2410.19599. <https://doi.org/10.48550/arXiv.2410.19599>
- Garaus, M., Weismayer, C., & Steiner, E. (2023). Is texture the new taste? The effect of sensory food descriptors on restaurant menus on visit intentions. *British Food Journal*, 125(10), 3817–3831. <https://doi.org/10.1108/BFJ-08-2022-0693>
- Girard, A., Lichters, M., Sarstedt, M., & Biswas, D. (2019). Short- and long-term effects of nonconsciously processed ambient scents in a servicescape: Findings from two field experiments. *Journal of Service Research*, 22(4), 440–455. <https://doi.org/10.1177/1094670519842333>
- Grewal, D., Baker, J., Levy, M., & Voss, G. B. (2003). The effects of wait expectations and store atmosphere evaluations on patronage intentions in service-intensive retail stores. *Journal of Retailing*, 79(4), 259–268. <https://doi.org/10.1016/j.jretai.2003.09.006>

- Guedes, D., Garrido, M. V., Lamy, E., Cavalheiro, B. P., & Prada, M. (2023). Crossmodal interactions between audition and taste: A systematic review and narrative synthesis. *Food Quality and Preference*, 107, 104856. <https://doi.org/10.1016/j.foodqual.2023.104856>
- Hermann, E., & Puntoni, S. (2024). Artificial intelligence and consumer behavior: From predictive to generative AI. *Journal of Business Research*, 180, 114720. <https://doi.org/10.1016/j.jbusres.2024.114720>
- Hollebeek, L. D., Sarstedt, M., Menidjel, C., Urbonavicius, S., & Dikcius, V. (2024). Theoretical rigor of customer experience scales: A systematic review and a roadmap for researchers. *Marketing Intelligence & Planning*, 42(8), 1464–1493. <https://doi.org/10.1108/MIP-06-2024-0413>
- Imschloss, M., & Kuehnl, C. (2017). Don't ignore the floor: Exploring multisensory atmospheric congruence between music and flooring in a retail environment. *Psychology & Marketing*, 34(10), 931–945. <https://doi.org/10.1002/mar.21033>
- Imschloss, M., & Kuehnl, C. (2019). Feel the music! Exploring the cross-modal correspondence between music and haptic perceptions of softness. *Journal of Retailing*, 95(4), 158–169. <https://doi.org/10.1016/j.jretai.2019.10.004>
- Ittamalla, R., & Srinivas Kumar, D. V. (2019). The impact of social cues on passengers' travel experience. *The Service Industries Journal*, 39(3–4), 299–318. <https://doi.org/10.1080/02642069.2018.1521389>
- Jin, C., Yoon, M., & Lee, J. (2019). The influence of brand color identity on brand association and loyalty. *Journal of Product & Brand Management*, 28(1), 50–62. <https://doi.org/10.1108/JPBM-09-2017-1587>
- Kabadayi, E. T., Aksoy, N. C., & Turkay, P. B. (2023). How does customer engagement value occur in restaurants? A stimulus-organism-response(S-O-R) perspective. *The Service Industries Journal*, 43(7–8), 497–524. <https://doi.org/10.1080/02642069.2022.2075350>
- Koffka, K. (1935). *Principles of gestalt psychology*. Routledge. <https://doi.org/10.4324/9781315009292>
- Krishna, A. (2012). An integrative review of sensory marketing: Engaging the senses to affect perception, judgment and behavior. *Journal of Consumer Psychology*, 22(3), 332–351. <https://doi.org/10.1016/j.jcps.2011.08.003>
- Labrecque, L. I., & Milne, G. R. (2012). Exciting red and competent blue: The importance of color in marketing. *Journal of the Academy of Marketing Science*, 40(5), 711–727. <https://doi.org/10.1007/s11747-010-0245-y>
- Li, J.-G. T., Kim, J.-O., & Lee, S. Y. (2009). An empirical examination of perceived retail crowding, emotions, and retail outcomes. *The Service Industries Journal*, 29(5), 635–652. <https://doi.org/10.1080/0264206902720121>
- Liu, A., Wang, X. L., Liu, F., Yao, C., & Deng, Z. (2018). Soundscape and its influence on tourist satisfaction. *The Service Industries Journal*, 38(3–4), 164–181. <https://doi.org/10.1080/02642069.2017.1382479>
- Madzharov, A., Ye, N., Morrin, M., & Block, L. (2018). The impact of coffee-like scent on expectations and performance. *Journal of Environmental Psychology*, 57, 83–86. <https://doi.org/10.1016/j.jenvp.2018.04.001>
- Mari, M., & Poggesi, S. (2013). Servicescape cues and customer behavior: A systematic literature review and research agenda. *The Service Industries Journal*, 33(2), 171–199. <https://doi.org/10.1080/02642069.2011.613934>
- Marjeh, R., Sucholutsky, I., Van Rijn, P., Jacoby, N., & Griffiths, T. L. (2024). Large language models predict human sensory judgments across six modalities. *Scientific Reports*, 14(1), 21445. <https://doi.org/10.1038/s41598-024-72071-1>
- Mattila, A. S., & Wirtz, J. (2001). Congruency of scent and music as a driver of in-store evaluations and behavior. *Journal of Retailing*, 77(2), 273–289. [https://doi.org/10.1016/S0022-4359\(01\)00042-2](https://doi.org/10.1016/S0022-4359(01)00042-2)
- McCoy, R. T., Yao, S., Friedman, D., Hardy, M., & Griffiths, T. L. (2023). *Embers of autoregression: Understanding large language models through the problem they are trained to solve*. arXiv Preprint. <https://doi.org/10.48550/arXiv.2309.13638>
- Milliman, R. E. (1982). Using background music to affect the behavior of supermarket shoppers. *Journal of Marketing*, 46(3), 86–91. <https://doi.org/10.1177/002224298204600313>

- Motoki, K., Spence, C., & Velasco, C. (2024). Colour/shape-taste correspondences across three languages in ChatGPT. *Cognition*, 253, 105936. <https://doi.org/10.1016/j.cognition.2024.105936>
- North, A. C. (2012). The effect of background music on the taste of wine: Music and wine. *British Journal of Psychology*, 103(3), 293–301. <https://doi.org/10.1111/j.2044-8295.2011.02072.x>
- Oakes, S., & North, A. C. (2008). Reviewing congruity effects in the service environment musicscape. *International Journal of Service Industry Management*, 19(1), 63–82. <https://doi.org/10.1108/09564230810855716>
- Salecha, A., Ireland, M. E., Subrahmanya, S., Sedoc, J., Ungar, L. H., & Eichstaedt, J. C. (2024). Large language models display human-like social desirability biases in big five personality surveys. *PNAS Nexus*, 3(12), Article pgae533. <https://doi.org/10.1093/pnasnexus/pgae533>
- 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 & Marketing*, 41(6), 1254–1270. <https://doi.org/10.1002/mar.21982>
- Sarstedt, M., Imschloss, M., & Adler, S. (2024). *Multisensory design of retail environments: Vision, sound, and scent*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-45678-8>
- Schouteten, J. J., Van Severen, A., Dull, D., De Steur, H., & Danner, L. (2024). Congruency of an eating environment influences product liking: A virtual reality study. *Food Quality and Preference*, 113, 105066. <https://doi.org/10.1016/j.foodqual.2023.105066>
- Siedenburg, K., & Saitis, C. (2023). *The language of sounds unheard: Exploring musical timbre semantics of large language models*. arXiv Preprint. <https://doi.org/10.48550/arXiv.2304.07830>
- Spangenberg, E. R., Grohmann, B., & Sprott, D. E. (2005). It's beginning to smell (and sound) a lot like Christmas: The interactive effects of ambient scent and music in a retail setting. *Journal of Business Research*, 58(11), 1583–1589. <https://doi.org/10.1016/j.jbusres.2004.09.005>
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, 73(4), 971–995. <https://doi.org/10.3758/s13414-010-0073-7>
- Spence, C. (2012). Managing sensory expectations concerning products and brands: Capitalizing on the potential of sound and shape symbolism. *Journal of Consumer Psychology*, 22(1), 37–54. <https://doi.org/10.1016/j.jcps.2011.09.004>
- Spence, C., & Levitan, C. A. (2021). Explaining crossmodal correspondences between colours and tastes. *I-Perception*, 12(3). <https://doi.org/10.1177/20416695211018223>
- Spence, C., Puccinelli, N. M., Grewal, D., & Roggeveen, A. L. (2014). Store atmospherics: A multisensory perspective. *Psychology & Marketing*, 31(7), 472–488. <https://doi.org/10.1002/mar.20709>
- Spence, C., Senkowski, D., & Röder, B. (2009). Crossmodal processing. *Experimental Brain Research*, 198(2), 107–111. <https://doi.org/10.1007/s00221-009-1973-4>
- Sugimori, E., & Kawasaki, Y. (2022). Cross-modal correspondence between visual information and taste perception of bitter foods and drinks. *Food Quality and Preference*, 98, 104539. <https://doi.org/10.1016/j.foodqual.2022.104539>
- Tili, A., Denden, M., Abed, M., & Huang, R. (2024). Artificial intelligence ethics in services: Are we paying attention to that? *The Service Industries Journal*, 44(15–16), 1093–1116. <https://doi.org/10.1080/02642069.2024.2369322>
- Uhrich, S., & Benkenstein, M. (2012). Physical and social atmospheric effects in hedonic service consumption: Customers' roles at sporting events. *The Service Industries Journal*, 32(11), 1741–1757. <https://doi.org/10.1080/02642069.2011.556190>
- Van Bergen, G., Ushijima, S., Kaneko, D., Dijksterhuis, G. B., De Wijk, R. A., & Vingerhoeds, M. H. (2022). What makes foods and flavours fit? Consumer perception of (un)usual product combinations. *Food Quality and Preference*, 102, 104680. <https://doi.org/10.1016/j.foodqual.2022.104680>
- Van Bergen, G., Zandstra, E. H., Kaneko, D., Dijksterhuis, G. B., & De Wijk, R. A. (2021). Sushi at the beach: Effects of congruent and incongruent immersive contexts on food evaluations. *Food Quality and Preference*, 91, 104193. <https://doi.org/10.1016/j.foodqual.2021.104193>
- Vatankhah, S., Bamshad, V., Arici, H. E., & Duan, Y. (2024). Ethical implementation of artificial intelligence in the service industries. *The Service Industries Journal*, 44(9–10), 661–685. <https://doi.org/10.1080/02642069.2024.2359077>

- Velasco, C., Woods, A. T., Deroy, O., & Spence, C. (2015). Hedonic mediation of the crossmodal correspondence between taste and shape. *Food Quality and Preference*, 41, 151–158. <https://doi.org/10.1016/j.foodqual.2014.11.010>
- Voss, K. E., Zablah, A. R., Huang, Y.-S. S., & Chakraborty, G. (2020). Conjunctionitis: A call for clarity in construct definitions. *European Journal of Marketing*, 54(5), 1147–1159. <https://doi.org/10.1108/EJM-07-2018-0470>
- Walters, J., Apter, M. J., & Svebak, S. (1982). Color preference, arousal, and the theory of psychological reversals. *Motivation and Emotion*, 6(3), 193–215. <https://doi.org/10.1007/BF00992245>
- Weng, C., Yuan, R., Ye, D., Huang, B., & Xun, J. (2024). Leveraging responsible artificial intelligence to enhance salespeople well-being and performance. *The Service Industries Journal*, 44(9–10), 735–765. <https://doi.org/10.1080/02642069.2024.2361291>
- Wexner, L. B. (1954). The degree to which colors (hues) are associated with mood-tones. *Journal of Applied Psychology*, 38(6), 432–435. <https://doi.org/10.1037/h0062181>
- Wright, A. (1995). *The beginner's guide to colour psychology*. Kyle Cathie Limited.
- Zellner, D., Geller, T., Lyons, S., Pyper, A., & Riaz, K. (2017). Ethnic congruence of music and food affects food selection but not liking. *Food Quality and Preference*, 56, 126–129. <https://doi.org/10.1016/j.foodqual.2016.10.004>
- Zlateva, N., Ivanov, S., & Fedoseev, V. (2024). Autonomous sensory meridian response in service experience: An exploratory study. *The Service Industries Journal*, 1–26. <https://doi.org/10.1080/02642069.2024.2319039>