



HAL
open science

Experiential Narratives in Marketing: A Comparison of Generative AI and Human Content

Yingting Wen, Sandra Laporte

► **To cite this version:**

Yingting Wen, Sandra Laporte. Experiential Narratives in Marketing: A Comparison of Generative AI and Human Content. *Journal of Public Policy and Marketing*, 2024, 10.1177/07439156241297973 . hal-04758305

HAL Id: hal-04758305

<https://em-lyon.hal.science/hal-04758305v1>

Submitted on 29 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Author Accepted Manuscript



**Experiential Narratives in Marketing: A Comparison of
Generative AI and Human Content**

Journal:	<i>Journal of Public Policy and Marketing</i>
Manuscript ID	JPPM-24-073.R2
Manuscript Type:	Special Issue Revised Submission
Topics and Methods:	Marketing and Society < Content Areas, Experimental Design < Methodological Areas, Content Analysis < Methodological Areas, Advertising Persuasion < Content Areas, Technology < Content Areas

SCHOLARONE™
Manuscripts

Author Accepted Manuscript

Experiential Narratives in Marketing: A Comparison of Generative AI and Human

Content

Yingting Wen

Assistant Professor of Marketing

EMLYON Business School

144 avenue Jean Jaurès, 69007 Lyon, France

wen@em-lyon.com

Sandra Laporte

Full Professor of Marketing

Toulouse School of Management

University Toulouse Capitole

2 rue du Doyen Gabriel Marty, 31042 Toulouse, France

sandra.laporte@tsm-education.fr

Acknowledgments: The authors would like to thank the editors, the guest editor, and the three anonymous reviewers for their invaluable feedback and advice that significantly improved this paper.

Author Accepted Manuscript

Experiential Narratives in Marketing: A Comparison of Generative AI and Human

Content

Abstract

As generative AI technologies advance, understanding their capability to emulate human-like experiences in marketing communication becomes crucial. This research examines whether generative AI can create experiential narratives that resonate with humans in terms of embodied cognition, affect, and lexical diversity. An automatic text analysis reveals that while reviews generated by ChatGPT 3.5 exhibit lower levels of embodied cognition and lexical diversity compared with reviews by human experts, they display more positive affect (Study 1A).

However, human raters struggle to notice these differences, rating half of the selected reviews from AI higher in embodied cognition and usefulness (Study 1B). Instances of hallucination in AI-generated content were detected by human raters. For social media posts, the more sophisticated ChatGPT 4 model demonstrates superior perceived lexical diversity and leads to higher purchase intentions in unbranded content compared with human copywriters (Study 2).

This paper evaluates the performance of large language models in generating experiential marketing narratives. The comparative studies reveal the models' strengths in presenting positive emotions and influencing purchase intent while identifying limitations in embodied cognition and lexical diversity compared to human-authored content. The findings have implications for marketers and policymakers in understanding generative AI's potential and risks in marketing.

Keywords: large language models, product description, embodied cognition, affect, lexical diversity, consumer perception

Author Accepted Manuscript

Ever since the growth of artificial intelligence (AI) during the 1990s, the marketing field has incorporated AI technology into various areas such as services, management of customer relationships, personalization, and targeting (Huang and Rust 2018; Libai et al. 2020; Ma and Sun 2020). Yet, the advent of ChatGPT marks another transformative moment in AI application for marketing (OpenAI 2022). Unlike other AI models that mainly focus on data analytics (Davenport et al. 2020), the generative AI model behind ChatGPT is designed to generate new content (e.g., texts, images, and videos) based on learning from large datasets in human knowledge (Cao et al. 2023). This model can understand and respond to human instructions, enabling the rapid production of large amounts of content. With their advanced language processing capabilities, generative AI models matched and sometimes outperformed humans in analogical reasoning tasks (Webb, Holyoak and Lu 2023). ChatGPT has successfully passed the US bar exam for licensed lawyers, accounting CPA exams for certified accountants, and even a medical licensing exam (Eulerich et al. 2023; Katz et al. 2023; Takagi et al. 2023). The revolutionary technology further creates new possibilities for marketing. According to surveys of marketers, marketing content creation is the most important use case for generative AI tools (Press 2023). For certain companies, ChatGPT is becoming a replacement for human copywriters (Verma and Vynck 2023).

Whereas generative AI technology seems extremely powerful, it is not infallible or without shortcomings. Recent research in computer science has tested the cognitive (Zhuang et al. 2023), emotional (Tak and Gratch 2023), and linguistic abilities (Digutsch and Kosinski 2023) of large language models (LLMs). Compared with humans, ChatGPT's semantic network is more based on semantic similarity (share of meaning in words, e.g., England-Britain) than on associative similarity (probability of co-occurrence in the language, e.g., Mountain-Top)

Author Accepted Manuscript

(Digutsch and Kosinski 2023). Large language models often fail in language tasks involving real-world knowledge from perceptual and social systems (Mahowald et al. 2023). For instance, large language models (LLMs) struggle to compare the sizes of objects like a mountain and a sofa because we rarely describe their actual dimensions explicitly. They also demonstrate inconsistent performance in inferring others' beliefs and intentions across various situations (Ullman 2023).

Even though generative AI models can mirror the human mind at a cognitive level, the question remains: can they replicate actual feelings and sensory experiences in consumption? Can generative AI create marketing content that resonates with human feelings and drives consumer behavior? This article examines the experiential narratives produced by generative AI models and compares their textual qualities with those created by humans. The traditional marketing approach that focuses on functional and economic values is gradually replaced by experiential marketing, which focuses on the sensory, affective, cognitive, and behavioral values in the consumption experience (Schmitt 1999). To enhance consumer connections to brands, experiential narratives in marketing leverage storytelling techniques to create immersive, engaging, and emotionally resonant consumer experiences (Escalas 2004). These narratives can manifest through various forms, such as customer testimonials/reviews, detailed descriptions of consumption experiences, and social media posts. For instance, ads that vividly depict the consumer's personal journey with the product can significantly improve brand attitudes and purchase intentions (Escalas, 2004). Similarly, detailed accounts of consumption experiences from a wine connoisseur discussing the wine's texture, color, and aroma can evoke sensory and emotional responses that strengthen consumer involvement (Mora and Moscarola 2010).

Author Accepted Manuscript

In theories of brand experience (Brakus, Schmitt and Zarantonello 2009), researchers highlight three core experiential dimensions for consumption. The sensory dimension refers to sensory perceptions experienced through consumers' senses—sight, sound, touch, taste, and smell. The emotional dimension relates to the feelings and emotions a brand elicits in consumers. The intellectual dimension involves the cognitive stimulation a brand provides that encourages consumers to think or engage creatively with the brand. In a similar vein, Schmitt (1999) introduces three strategic experiential modules in experiential marketing: sensory experiences (SENSE), affective experiences (FEEL), and creative cognitive experiences (THINK). Building on these theories, we analyze experiential narratives in marketing using three types of textual cues corresponding to these three dimensions. We use embodied cognition to measure the extent to which bodily experiences and physical interactions are involved in describing sensory experiences. We examine affect, as presented in texts through emotions and tones, to explore the inclusion of affective experiences. Finally, we assess lexical diversity to determine whether the narratives engage consumers creatively with rich and varied language. Table 1 summarizes the key terms used in the current research. Our study adopts a multi-method approach, combining objective analysis through automated textual analysis with subjective analysis based on human raters' perceptions. We study two types of marketing communication that leverage experiential narratives in food and beverage consumption: product experience reviews and social media posts. These types of marketing content represent evolving channels of online word of mouth (WOM) in marketing communication that drive product sales (Kumar et al. 2016; Zhu and Zhang 2010). The use of generative AI to summarize, edit, or generate customer reviews signifies a transformative shift in marketing communication strategies with potential ethical concerns. Recently, the Amazon platform started providing AI-generated customer “highlights,”

Author Accepted Manuscript

which are short paragraphs that summarize common themes, product features, and customer sentiment based on existing customer reviews (Schermerhorn 2023). However, generative AI can also be used to generate fake customer reviews or other experiential narratives from scratch, which will be the focus of this research. The growing use of generative AI in marketing content creation—both ethical and unethical—demands investigation into its potential consequences. Larger language models like ChatGPT did not learn bodily experiences through direct interaction with the physical world. Instead, they summarize human sensory experiences by processing and interpreting human languages from extensive datasets. This paper's core objective is to examine the textual qualities of content created by generative AI in sensory consumption experiences and enhance our understanding of generative AI's role in content generation within marketing communication. While existing research on generative AI content predominantly utilizes objective and technical metrics for comparison (Digutsch and Kosinski 2023), this investigation takes a different approach. It complements the objective analysis provided by automated textual analysis with assessments of AI content from the consumers' perspective, exploring their perceptions, comprehension, and emotional responses to AI-generated product descriptions.

Our analysis reveals that, compared with human experts, ChatGPT 3.5 can exhibit lower levels of embodied cognition and lexical diversity. Interestingly, human raters often fail to detect these differences but can find evidence of hallucinations (i.e., instances where a language model generates factually incorrect texts, see Table 1) in AI-generated content. In contrast, the more advanced ChatGPT 4 model can match or even surpass human copywriters in perceived lexical diversity and effectiveness in driving purchase intention. These findings underscore the potential of generative AI in creating marketing narratives while highlighting the risk of producing inaccurate content. To the best of our knowledge, this is the first study to examine consumer

Author Accepted Manuscript

perceptions of experiential narratives in marketing produced by generative AI. Although large language models (LLMs) have been extensively studied in computer science for their performance on cognitive and language tasks (Srivastava et al. 2022) , there is scant research on their application in marketing content creation and subsequent consumer responses. This study bridges that gap by evaluating the effectiveness of LLMs in generating marketing narratives and understanding how these AI-generated messages resonate with consumers. Our findings offer marketers strategic insights for AI integration in marketing communication while informing policymakers to develop ethical guidelines for generative AI in marketing.

Peer Review Version

Author Accepted Manuscript

Table 1. Key Terminology and Related Literature

Terminology	Definition	Research in Marketing	Research in AI	This Paper
Embodied cognition	Cognitive processes that are deeply rooted in the body's interactions with the physical world. For instance, people understand and remember geometry concepts when they physically interact with objects in those shapes (Wilson 2002).	Sensory experiences and mental simulation from embodied cognition drive consumer preferences and purchase intentions (Kwon and Adaval 2018; Elder and Krishna 2011; Krishna and Schwarz 2014; Cascio Rizzo et al. 2023).	LLMs lack physical bodies essential for sensory experiences, leading to less nuanced descriptions (Xu et al. 2023). ChatGPT 4, with visual data training, shows enhanced performance in mental visualization (Xu et al. 2023).	Whether generative AI can produce experiential narratives with high levels of embodied cognition and how consumers react to them.
Affect	Experiences of positive or negative feelings that include both the emotions expressed and the tones used in communications (Downes 2000; Russell 2003).	Affect influences advertising effectiveness and viral content (Poels and Dewitte 2019; Lovett, Peres, and Shachar 2013; Berger and Milkman 2012; Rocklage and Fazio 2020; Eckler and Bolls 2011).	AI models can create emotionally resonant content (Ray 2023; Mishra et al. 2023; Lynch et al. 2023), but struggle with emotion forecasting and empathy (Tak and Gratch 2023; De Freitas et al. 2023;	The valence and extremity of emotions and tones in experiential narratives created by generative AI.

Author Accepted Manuscript

			Liu-Thompkins, Okazaki, and Li 2022).	
Lexical diversity	The range of unique words used in a text that reflects the richness and variability of language (McCarthy and Jarvis 2010).	Lexical diversity affects engagement in podcasts and consumer reviews (Liu et al. 2023; Liu, Xie, and Zhang 2019).	ChatGPT 4 exhibits high lexical diversity but reduced readability (Herbold et al. 2023; AlAfnan and MohdZuki 2023).	The lexical diversity of experiential narratives from generative AI and consumers' perceptions.
Hallucination	In the AI context, refers to situations where AI models confidently provide responses that appear plausible but are factually incorrect or nonsensical when evaluated against common knowledge (Maleki, Padmanabhan and Dutta 2024).	-	AI models produce hallucinations due to issues with both training datasets and model architectures (Dziri et al. 2022).	The existence of hallucinations in experiential narratives produced by generative AI and consumers' reactions.

Author Accepted Manuscript

Theoretical Background

Based on the three key modules of experiential marketing—sensory experiences, affective experiences, and creative cognitive experiences (Schmitt 1999)—this research focuses on three crucial textual dimensions for narrating sensory consumption experiences. The first dimension, embodied cognition, is the ability to articulate and understand sensory experiences derived from physical interactions with the environment. Affect, the second dimension, demonstrates the capacity to comprehend and convey emotions through content or tone. Finally, to generate creative cognitive experiences, it is essential to use unique and varied descriptions to avoid repetitive experiential narratives. The content generated from consumption experiences should display substantial lexical diversity, which may present a challenge for generative AI. Together, these dimensions provide a comprehensive framework for analyzing the depth and complexity of experiential narratives in marketing. The primary goal of this research is to empirically compare the performance of two generations of generative AI (ChatGPT 3.5 and ChatGPT 4) with human-generated content. Given the exploratory nature of this investigation and the scarcity of prior research comparing these AI models to human performance, we do not formulate specific hypotheses. Instead, we aim to uncover nuanced insights into the capabilities and limitations of these AI models, thereby contributing to a deeper understanding of their potential applications and impact across various contexts.

Embodied Cognition

While traditional cognitive theories view information processing as computations isolated from bodily experiences, embodied cognition emphasizes that cognitive processes are deeply

Author Accepted Manuscript

intertwined with bodily interactions with the environment and highlights the importance of sensorimotor experiences in our cognitive abilities (Foglia and Wilson 2013). Bodily experiences are physical sensations such as touching or seeing that influence our thoughts. As we engage with our environment by navigating a room or handling objects, bodily interactions further shape our understanding and memory. Additionally, sensorimotor experiences that combine sensory inputs (what we see, hear, feel) with motor actions (movements) form the foundation for language comprehension. For instance, the phrase “financial burden is weighing him down” makes sense because we have all physically carried something heavy and felt the weight. This physical sensation helps us understand the emotional burden of financial stress. Evidence from neuroscience shows that verb processing is associated with brain regions for motor control functions, just as color naming is connected to brain areas for visual perception (Anderson, Richardson and Chemero 2012). Patients with motor system damage also have difficulty understanding action verbs like “running,” “kicking,” and “grasping” (Cardona et al. 2014). In marketing, sensorimotor experiences also play a key role in consumers’ preferences for options (Kwon and Adaval 2018) and purchase intention (Elder and Krishna 2011). Findings from the field of sensory marketing further highlight that sensory experiences driven by physical stimuli and mental stimulation have an important impact on marketing outcomes (Krishna and Schwarz 2014). Influencers who use more sensory language in their content also attain more engagement and sales (Cascio Rizzo et al. 2023).

Large language models (LLMs) do not own a physical body that is essential for sensory experiences. As a result, LLMs rely on vast volumes of training data to learn from those experiences. This lack of embodiment could be a limitation for LLMs: their abilities to describe sensory experiences may be less nuanced or detailed than that of humans, and their descriptions

Author Accepted Manuscript

might lack the depth and authenticity that humans possess. Xu and colleagues (2023) studied how ChatGPT 3.5 and 4 performed compared to humans when rating lexical presentations for embodied experience (e.g., how much do you experience the concept of flower by touching/smelling/listening). They show that both LLMs can match human ratings in more abstract domains like emotions, but their ratings deviate from human ratings in sensory (related to senses like sight, sound, and touch) and motor (related to physical actions and movements) domains (Xu et al. 2023). Due to a lack of embodied experience, GPT 3.5 and GPT 4 fail to consistently align with human ratings in sensory dimensions such as visual, olfactory, and tactile representations of flowers (correlations $<.60$). However, as the more advanced model incorporating visual data training, ChatGPT 4 demonstrate enhanced performance in mental visualization (Xu et al. 2023). For hedonic products (e.g., food, cosmetics, art, etc.), their marketing communications (e.g., reviews and ad copywriting) require accurate portrayals of sensory experiences. That raises questions on whether generative AI can have comparable performance in creating embodied cognition (or anything close to embodiment) within marketing communications for these types of products.

Affect and Marketing Communication

Affect has always been crucial in marketing communications, including advertising (Poels and Dewitte 2019) and word-of-mouth (WOM) (Lovett, Peres and Shachar 2013). By studying viral online content, Berger and Milkman (2012) find that content with high-arousal emotions, whether positive or negative, is more likely to go viral than content with low-arousal emotions. Yet, for consumer reviews, the effect of positive emotions mainly applies to hedonic products, as these emotions can induce mistrust in reviews of utilitarian products (Rocklage and Fazio 2020). Besides the emotions expressed in marketing communications, the tone used in the messages is

Author Accepted Manuscript

also a type of affect that influences marketing outcomes. For instance, research by Eckler and Bolls (2011) demonstrates that video advertisements with a more positive tone increase positive attitudes toward the advertisements and sharing intention.

Despite not having direct experiences with emotions, advanced Generative AI models like ChatGPT can create emotionally resonant content without possessing actual emotional intelligence (Ray 2023). They can generate human-like and appropriate emotional expressions in conversations (Mishra et al. 2023) and narrative generation (Lynch et al. 2023). Nevertheless, more recent research uncovers that generative AI models fail to forecast emotion intensities and provide coping responses (Tak and Gratch 2023). Chatbots using generative AI models also struggle to address consumers' negative emotions and mental health issues appropriately (De Freitas et al. 2023). Present AI agents still fall short in empathy for customers' affective and social experiences (Liu-Thompkins, Okazaki and Li 2022). Thus, using generative AI for content creation in marketing communications entails potential risks associated with the absence or misuse of emotions.

Lexical Diversity

Prior studies have shown that lexical diversity, which refers to the range of unique words within a text (McCarthy and Jarvis 2010), can affect message effectiveness for high-status speakers. Podcasts with a moderate level of lexical diversity attain the highest level of listener engagement (Liu et al. 2023). However, the findings are more nuanced in marketing. For consumer reviews, language style matching between a reviewer and readers is more important than specific lexical features in the perceived helpfulness of the review (Liu, Xie and Zhang 2019). On the other hand, products like music need repetitive lyrics with low lexical diversity to generate higher processing fluency and market success (Nunes, Ordanini and Valsesia 2015).

Author Accepted Manuscript

While writing from ChatGPT 3 exhibits less linguistic diversity than student essays (as rated by human experts), ChatGPT 4 can already use more complex and diverse vocabulary compared to human authors (Herbold et al. 2023). Similarly, in business correspondence writing, generative AI models like ChatGPT 4 can achieve high lexical diversity, albeit with reduced readability (Alafnan and MohdZuki 2023). One exception occurs in the domain of news articles (Tseng, Verberne and van der Putten 2023), where human comments consistently demonstrate higher lexical diversity than ChatGPT 3.5. Based on previous findings, generative AI should be able to match or even surpass humans in lexical diversity for simple marketing copywriting tasks. Yet, it would be interesting to explore if this holds across different contexts and audiences and how it impacts consumers' evaluations and engagement.

Overview of Studies

The current research comprises three empirical studies that investigate how generative AI and humans differ in their experiential narratives (see Table 2 for a summary). Focusing on sensory experiences of two product categories (whisky and chocolate), the studies compare content from generative AI and humans on their embodied cognition, affect, and lexical diversity with a multi-method approach. Study 1A uses automatic text analysis as an objective lens to compare whisky reviews written by human experts against those generated by ChatGPT 3.5 for identical products. It demonstrates that human experts exhibit greater embodied cognition and lexical diversity, while ChatGPT 3.5 displays more positive tones and emotions. Study 1B extends the investigation of Study 1A by using human raters instead of automatic text analysis tools to evaluate a sample of reviews from Study 1A. Study 2 is a comparative analysis of social media

Author Accepted Manuscript

content related to chocolate products on Instagram. The posts used as stimuli include branded content (posts promoting specific products or brands) and unbranded content (posts that do not promote specific products or brands) written by either verified accounts or ChatGPT 4.0, a more advanced and multimodal model. We distinguish these two types of content based on past research showing that branded and unbranded content in social media marketing engage and persuade audiences through different mechanisms (Darmawan and Huh 2022; Holiday et al. 2023). Human judges evaluate these posts on embodied cognition, affect, and lexical diversity, as well as their potential downstream consequences, such as audience engagement and purchase intention. The data and stimuli of the studies are available on OSF at:

https://osf.io/gc7xz/?view_only=c8cb695057204d278aa2826feb59a32d.

Table 2. Summary of Studies

Studies	Stimuli	Purpose	Method	Key Measures
1A	Reviews of product experience (whisky)	Compare linguistic characteristics of generative AI and human reviews	Automatic text analysis (on 222 whisky products)	Embodied cognition; affect; lexical diversity
1B	Reviews of product experience (whisky)	Consumer perceptions and reactions to human expert reviews	Human raters ($N = 205$)	Embodied cognition; affect; lexical diversity; authenticity; expertise; usefulness

Author Accepted Manuscript

and generative

AI content

2	Social media posts (chocolate products)	Compare human and generative AI content in a different context	Human raters ($N = 207$)	Embodied cognition; affect; lexical diversity; engagement; purchase intention
---	---	---	-------------------------------	--

Study 1A

Study 1A provides a preliminary test of the linguistic characteristics in texts produced by human experts and a widely used generative AI model (ChatGPT). Using automatic text analysis, this study compares two samples of whisky reviews for the same group of products to investigate differences in embodied cognition, affect, and lexical diversity.

Data and Procedure

Whisky reviews were collected from two sources: human expert critiques in Whisky Advocate and outputs by Open AI's ChatGPT 3.5 (<https://chat.openai.com/>), an AI model free to the public. Whisky Advocate is a leading whisky magazine featuring whisky news and reviews from renowned experts. An archive of 2,247 reviews written by expert critics is accessible at <https://www.kaggle.com/datasets/koki25ando/22000-scotch-whisky-reviews>. Considering the output constraints of ChatGPT 3.5, a stratified sample of 222 reviews (one review per whisky product) was generated. We first broke down the reviews by five whisky categories (single malt

Author Accepted Manuscript

scotch, blended scotch whisky, blended malt scotch whisky, single grain whisky, and grain scotch whisky) and further categorized them based on two key product attributes: product price and review point (i.e., the overall evaluation of the reviewed product on a 100-point scale). We defined the low (≤ 25 th percentile), medium (above 25th percentile but ≤ 75 th percentile), and high (above 75th percentile) ranges for both attributes. This classification resulted in 5 (product categories) \times 3 (price ranges) \times 3 (review point ranges) = 45 distinct groups. To ensure that each group was proportionally represented in our sample, we randomly selected 10% of the reviews from each group to create the final sample (see Web Appendix A for summary statistics). The final sample encompassed reviews from all whisky categories, price ranges, and review point ranges, with lengths ranging from 20 to 200 words and an overall average of 72.45 words. Next, we retained the full names of 222 whisky products reviewed in this sample and created new corresponding whisky reviews using ChatGPT 3.5. ChatGPT 3.5 generated only the review texts, while all other product attributes remained unchanged from the Whisky Advocate dataset. To generate the review texts for the ChatGPT sample, we utilized the following standardized prompt:

As a writer for Whisky Advocate, a magazine dedicated to whisky reviews, I am seeking highly detailed reviews for a specific selection of whiskies. Each review should be between 20 and 220 words long (with an average of 70 words), ensuring a comprehensive and precise description of each whisky. Please continue generating reviews for each whisky on the list until the token limit is reached.

Here is the list line by line:

[a list of whisky names]

Author Accepted Manuscript

Based on the token limit of 4096 for ChatGPT 3.5, which refers to the maximum number of tokens (units of text) the model can process at once (including both input and output), we input approximately 20 whisky names into the list each time until all the products were included. This process ensured that the model could generate detailed reviews without exceeding its capacity. To ensure the length of the reviews from ChatGPT matched those of the human experts sample, longer reviews were requested using the prompt “Please generate more reviews with more than 70 words” if more than half of the initial reviews had word counts under 70. This process resulted in 222 ChatGPT-generated reviews that matched the list of products from the stratified sample of human expert reviews. The average length of ChatGPT-generated reviews was 71.27 words. A Mann-Whitney U test was conducted to compare word counts in both review samples (due to violations of the normality and homogeneity of variances assumptions). The result indicated no statistically significant difference in word counts between the two samples, $U = 25,453, p = .55$.

The linguistic characteristics of the review samples were analyzed using the Linguistic Inquiry and Word Count (LIWC) program (Boyd et al. 2022). LIWC provided text analysis outputs on various dimensions and sub-dimensions based on dictionaries related to psychological and social constructs. To measure embodied cognition, LIWC scores on cognitive processes (including dictionary words related to processes like insight, causation, and differentiation) and perception (including words on categories like attention, motion, visual, and auditory) were summed. This approach guaranteed that the sensory language incorporated cognitive reflection. For affect measures in LIWC, we retrieved dimensions at two levels: affect as a general concept at the first-level dimension, and sub-dimensions such as positive/negative tones and positive/negative emotions. Finally, for lexical diversity, the analysis included calculating the

Author Accepted Manuscript

overall type-token ratio (TTR), the mean segmental type-token ratio (MSTTR) (Johnson 1944) across two samples, and a comparison of word distribution (more details are provided in the relevant result section below).

Results

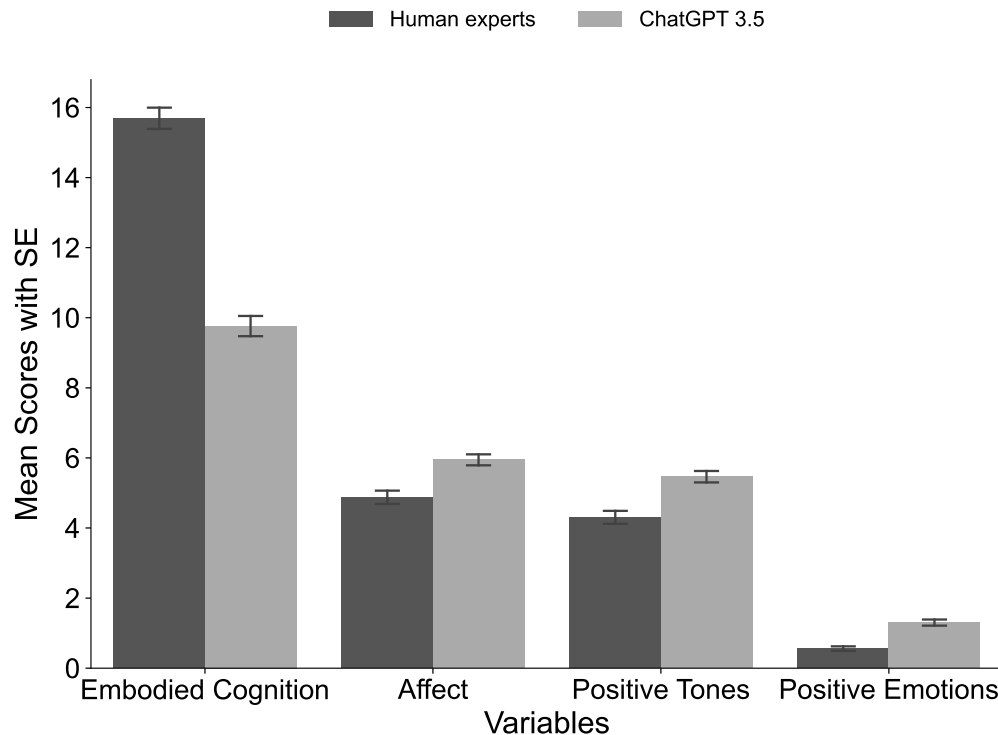
Shapiro-Wilk tests on reviews from human experts and ChatGPT revealed significant non-normality in most linguistic variables. For human expert reviews, the distributions for embodied cognition ($W = 0.99, p = .21$), affect ($W = 0.972, p < .001$), positive tones ($W = 0.96, p < .001$), negative tones ($W = 0.49, p < .001$), positive emotions ($W = 0.65, p < .001$), and negative emotions ($W = 0.23, p < .001$) significantly deviated from normality. Similarly, for ChatGPT reviews, significant departures from normal distribution were observed for embodied cognition ($W = 0.98, p = .005$), affect ($W = 0.99, p = .047$), positive tones ($W = 0.98, p = .012$), negative tones ($W = 0.64, p < .001$), and positive emotions ($W = 0.86, p < .001$), with only zeros in values for negative emotions. Given these violations of assumptions necessary for parametric testing, only non-parametric methods, specifically the Mann-Whitney U tests, were employed for the result analysis.

Embodied cognition

As mentioned before, embodied cognition was calculated by summing the values of the cognitive process and perception variables from LIWC. The Mann-Whitney U test revealed a significant difference in the level of embodied cognition between the two samples, $U = 40,939, p < .001$. ChatGPT reviews ($M = 9.76, SD = 4.32$) had a significantly lower level of embodied cognition than human expert reviews ($M = 15.69, SD = 4.51$, see Figure 1).

Author Accepted Manuscript

Figure 1. Comparative Analysis of Linguistic Features in Study 1A



Note: Error bars represent standard errors.

Affect

There were also significant differences in affect measures between ChatGPT and human expert reviews (see Figure 1). In LIWC, “affect” measures the presence of any affective content in texts. It includes two sub-measures: “emotions,” which assesses the emotional expressions in the content, and “tones,” which analyzes the positive or negative sentiment in the language.

ChatGPT reviews demonstrated a significantly higher level of affect ($M = 5.95$, $SD = 2.33$) than human expert reviews ($M = 4.88$, $SD = 2.82$), $U = 18,138.5$, $p < .001$. This was mainly because ChatGPT displayed significantly more positive tones ($M = 5.46$, $SD = 2.45$, $U = 17,585.5$, $p < .001$) and positive emotions ($M = 1.30$, $SD = 1.31$, $U = 16,502$, $p < .001$) compared with human experts (positive tones: $M = 4.31$, $SD = 2.75$; positive emotions: $M = 0.56$, $SD = 0.95$).

Lexical diversity

Author Accepted Manuscript

The Type-Token Ratio (TTR) and the Mean Segmental Type-Token Ratio (MSTTR) are indices of lexical diversity in linguistic research (Johnson 1944). TTR is a measure of vocabulary variety, calculated as the ratio of different words to the total number of words (tokens) in a text. This metric is particularly useful for evaluating the richness and variability of language use and is sensitive to the length of the texts. The MSSTR overcomes the length-related bias by calculating the TTR over fixed-sized segments of text (e.g., every 1000 words). It provides a more consistent measure of lexical diversity across texts of varying lengths. To calculate the overall Type-Token Ratio (TTR) and Mean Segmental Type-Token Ratio (MSTTR) per 1000 words of both samples, reviews from all whisky products were consolidated per sample for a comprehensive evaluation of lexical diversity. Natural language processing techniques were used to prepare the text data before calculating the TTR and MSTTR. The preprocessing included tokenization, which broke down text into individual units like words or phrases. The next step involved stop word removal, which excluded common words such as “the” and “is” that don’t add much meaning. Finally, lemmatization was applied to simplify words to their base forms, so words like “running,” “ran,” and “runs” are all reduced to “run.” These steps helped standardize the text for more accurate analysis. The TTR and MSTTR were then calculated for each corpus of reviews, resulting in two TTR and MSTTR values: one for the human review sample and one for the ChatGPT 3.5 sample. The results revealed that reviews from the human experts had a higher TTR (0.26) and MSTTR (0.59) compared with ChatGPT 3.5 (TTR: 0.10; MSTTR: 0.29), suggesting human experts used a wider variety of vocabulary across texts. To provide a statistical comparison of lexical diversity, we conducted a Chi-square test on the distribution of the most frequent words between the two samples, as discussed in the following analysis.

Author Accepted Manuscript

Using identical data preprocessing procedures (i.e., tokenization, stop word removal, and lemmatization), the top 20 words were identified for human expert and ChatGPT reviews (see Figure 2). There was a shared core vocabulary centered on essential aspects of the whisky-tasting experience, such as words like finish and palate. Both human experts and ChatGPT 3.5 emphasize the common taste (fruit, spice) and aroma (nose) of the whisky, which are fundamental components in reviews. Nonetheless, there were notable differences between the two review samples. Human experts used more terms related to aging (year and old) and cask types (cask). They also employed a broader selection of taste descriptors, such as toffee, chocolate, and honey. These choices of words were indicative of more detailed and nuanced observations, possibly drawing from richer experiences.

Figure 2. Frequency of Top 20 Words in Whisky Reviews: Human Experts vs. ChatGPT 3.5

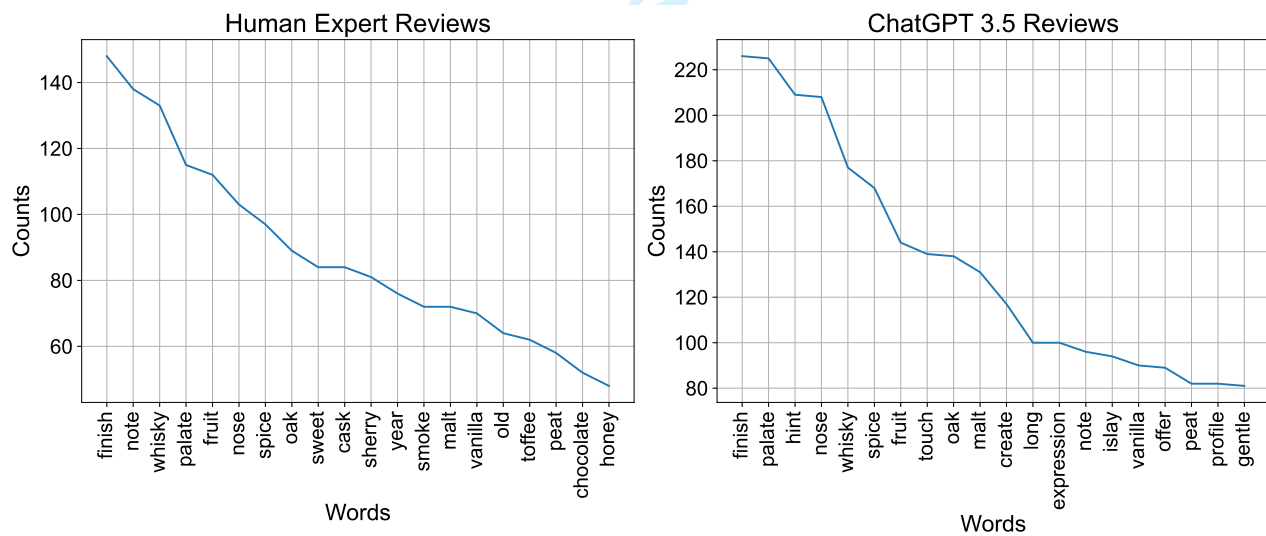


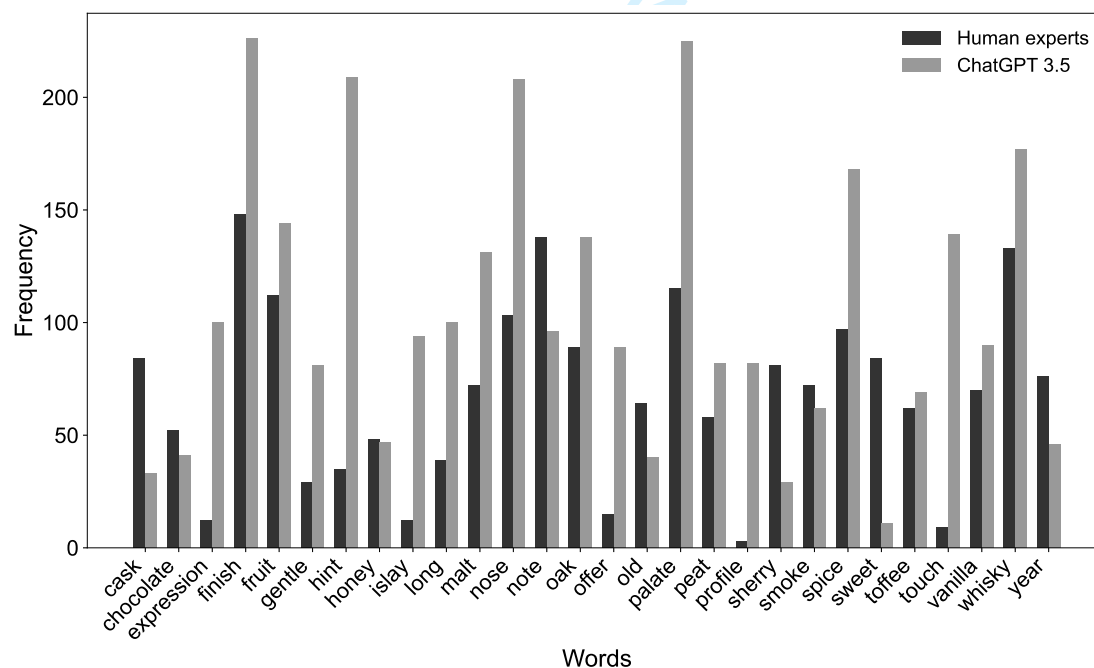
Figure 3 presents a side-by-side analysis by combining all top words and their corresponding frequencies from each sample. A chi-square test that compared the top words distributions further revealed a significant difference between the two samples ($\chi^2 = 629.71$, $p < .001$). The figure illustrated that ChatGPT 3.5 tended to overuse its top words with a much

Author Accepted Manuscript

higher frequency compared with human experts, implying that the AI model relied on certain words more heavily due to its training data patterns.

Additionally, there were distinct gaps in the distribution of specific words. For example, human experts utilized words like chocolate, sweet, and sherry more often, which were less prevalent in the AI-generated texts. On the other hand, words like hint, finish, and expression were used more frequently by ChatGPT 3.5. These types of words are more functional as they describe the overall impression rather than specific sensory details. Therefore, the AI model might prioritize general descriptors that apply to a wider range of experiences, while human experts delve into more detailed sensory descriptions that convey the specific nuances of each whisky.

Figure 3. Word Frequency Distribution by Data Source



Discussion

Author Accepted Manuscript

With automatic text analysis, Study 1A demonstrates initial evidence of how human experts and generative AI differ when describing consumption experiences. The result uncovers that reviews written by human experts have higher levels of embodied cognition and lexical diversity.

Compared with outputs from ChatGPT 3.5, reviews from human experts reflect richer bodily interactions from their unique physical experiences. They also use more diverse and specific vocabulary in their reviews. The top words analysis revealed a disproportionate representation of certain words in content from generative AI. This overrepresentation could stem from biases present in the training data, including stereotypes, which the models further amplified by reinforcing associations among word tokens (Liang et al. 2021). However, human experts display much less positive tones and emotions, which could be due to the emotional numbness (i.e., significantly lower emotional reactions to hedonic experiences) of experts (Rocklage, Rucker and Nordgren 2021). Another possible explanation could be the strategies used by developers to fine-tune generative AI models that aim to improve their interactions with humans, as evidenced by the more positive emotions and tones.

Study 1B

Although automatic text analysis tools can identify differences between reviews from human experts and generative AI, can humans perceive those differences without relying on extensive vocabulary dictionaries and quantitative algorithms? Study 1B uses a selection of reviews from Study 1A and asks human raters to assess them on the same variables. The goal is to examine whether consumers, as the audience of these reviews, can detect tangible differences between the content written by human experts and that generated by ChatGPT 3.5.

Author Accepted Manuscript

Data and Procedure

A sample of whisky reviews was generated by randomly selecting four human expert reviews, each representing different levels of embodied cognition in four quartiles (below the 25th percentile, between the 25th and 50th percentiles, between the 50th and 75th percentiles, and above the 75th percentile) from reviews in Study 1A. Subsequently, the corresponding ChatGPT reviews for the same whisky products were identified. The final stimuli consisted of four reviews from human experts and four ChatGPT reviews for the same whisky products.

Two hundred five alcohol consumers (49.27% female, 50.73% male; $M_{age} = 32.43$, $SD_{age} = 9.70$) were recruited from Prolific for an online study. All participants passed an attention check question, so there was no exclusion from the sample. They were randomly assigned to one of the two conditions: human expert reviews or ChatGPT reviews. The order of the reviews (four in total for each condition) was randomized. After reading a review, subjects evaluated it on several dimensions in 7-point Likert scales. Embodied cognition was measured with a three-item scale developed based on prior work from Barsalou (2008) (“The review makes me feel as if I can almost smell and taste the product,” “When I read the review about the product, I vividly imagined its taste and texture,” “I felt a strong physical craving for the product after reading the review”); $\alpha > .85$ for all reviews). Positive emotions were assessed with the item “How positive are the emotions expressed in the review?” (1 = Not positive at all, 7 = Very positive). Lexical diversity was evaluated using items created based on Jarvis’s (2013) definition of the construct: “How would you describe the choice of words in the review?” (1 = With low diversity/Very repetitive/Very limited, 7 = With high diversity/Very varied/Very rich; $\alpha > .80$ for all reviews). For exploratory purposes, the perceptions of authenticity (Beverland and Farrelly 2010; $\alpha > .87$ for all reviews; “The content of the review is realistic/truthful/authentic,” 1 = Strongly disagree,

Author Accepted Manuscript

7 = Strongly agree), level of expertise (1 = Low level of expertise, 7 = High level of expertise), and usefulness (1 = Not useful at all, 7 = Very useful) of the review were also measured. In the end, participants left their thoughts and comments on the study in an optional open-ended question.

Results

The distributions of most key variables (11 out of 12) deviated from a normal distribution. Hence, non-parametric Mann-Whitney U tests were used for group comparisons. Table 3 summarizes the descriptive statistics and results of Mann-Whitney U tests for key variables across two conditions.

Table 3. Descriptive Statistics of Key Variables in Study 1B

Reviews	Variables					
	Embodied Cognition		Positive Emotions		Lexical Diversity	
	Human	ChatGPT	Human	ChatGPT	Human	ChatGPT
	Experts	3.5	Experts	3.5	Experts	3.5
Review 1	4.94	5.12	5.38	5.71	5.35	5.27
	(1.53)	(1.23)	(1.19)	(1.09)	(1.17)	(1.05)
MWU	U = 5481, $p = .59$		U = 6108, $p = .036^*$		U = 4919, $p = .43$	
Review 2	4.73	5.04	5.39	5.76	4.99	5.31
	(1.38)	(1.22)	(1.28)	(0.97)	(1.28)	(1.02)
MWU	U = 5843, $p = .16$		U = 6000, $p = .068^†$		U = 5952, $p = .10$	
Review 3	4.11	4.56	5.06	5.34	4.86	5.12
	(1.44)	(1.34)	(1.33)	(1.28)	(1.19)	(1.14)

Author Accepted Manuscript

MWU	U = 6258, $p = .017^*$		U = 5900, $p = .12$		U = 5834, $p = .17$	
Review 4	4.32	4.77	5.48	5.56	5.16	5.42
	(1.67)	(1.42)	(1.43)	(1.32)	(1.33)	(1.20)
MWU	U = 6095, $p = .047^*$		U = 5843, $p = .81$		U = 5843, $p = .20$	

Note: Means with standard deviations in parentheses. MWU: Mann-Whitney U tests. † $p < .10$. * $p < .05$.

Embodied cognition

Human raters observed no significant differences in embodied cognition between reviews by human experts and ChatGPT for Reviews 1 and 2 (see Table 3). Contrary to the embodied cognition scores from LIWC, they evaluated ChatGPT reviews as having higher levels of embodied cognition for Reviews 3 and 4 (see Web Appendix B for the actual reviews and LIWC scores of the sample).

Positive emotions

The analysis of affect in Study 1B mainly focused on positive emotions because LIWC scores on positive tones were similar between human experts and ChatGPT reviews ($M_{human\ experts} = 5.51$, $M_{ChatGPT} = 5.61$). According to participants' ratings, ChatGPT showed more positive emotions in Reviews 1 and 2 (see Table 3). For Reviews 3 and 4, there were no significant differences in positive emotions. The participants' evaluations were thus consistent with LIWC scores for half the sample, which indicated that compared with human experts, generative AI like ChatGPT often showed more positive emotions in product reviews, even to human readers.

Lexical diversity

Author Accepted Manuscript

No significant difference in lexical diversity was found between the two groups of reviews (see Table 3). This outcome could be due to human raters' limited exposure to product reviews.

Unlike automatic text analysis tools that utilize extensive dictionaries and evaluate reviews with a statistical approach as seen in Study 1A, human raters encountered difficulties in assessing vocabulary diversity and identifying the overuse of certain word groups when only a limited number of reviews were presented to them. Although human experts showed objectively a greater overall lexical diversity in their reviews in Study 1A, this advantage did not significantly impact participants in Study 1B who read only four product reviews. This smaller sample might give them few opportunities to notice the recurrence of some words and the lower lexical diversity. These conditions are likely to correspond to the browsing behavior of online reviews in real life, where consumers will only read a reasonable sample of reviews.

Other variables

There was no significant difference between the two conditions in authenticity, expertise, and usefulness for Reviews 1-2. For Review 3, the ChatGPT-generated review showed marginally higher levels of expertise and usefulness (see Table 4). For Review 4, The ChatGPT review was judged to have significantly higher levels of authenticity, expertise, and usefulness (see Figure 4).

Table 4. Descriptive Statistics of Other Variables in Study 1B

Reviews		Variables			
		Authenticity		Expertise	
Human	ChatGPT	Human	ChatGPT	Human	ChatGPT
Experts	3.5	Experts	3.5	Experts	3.5

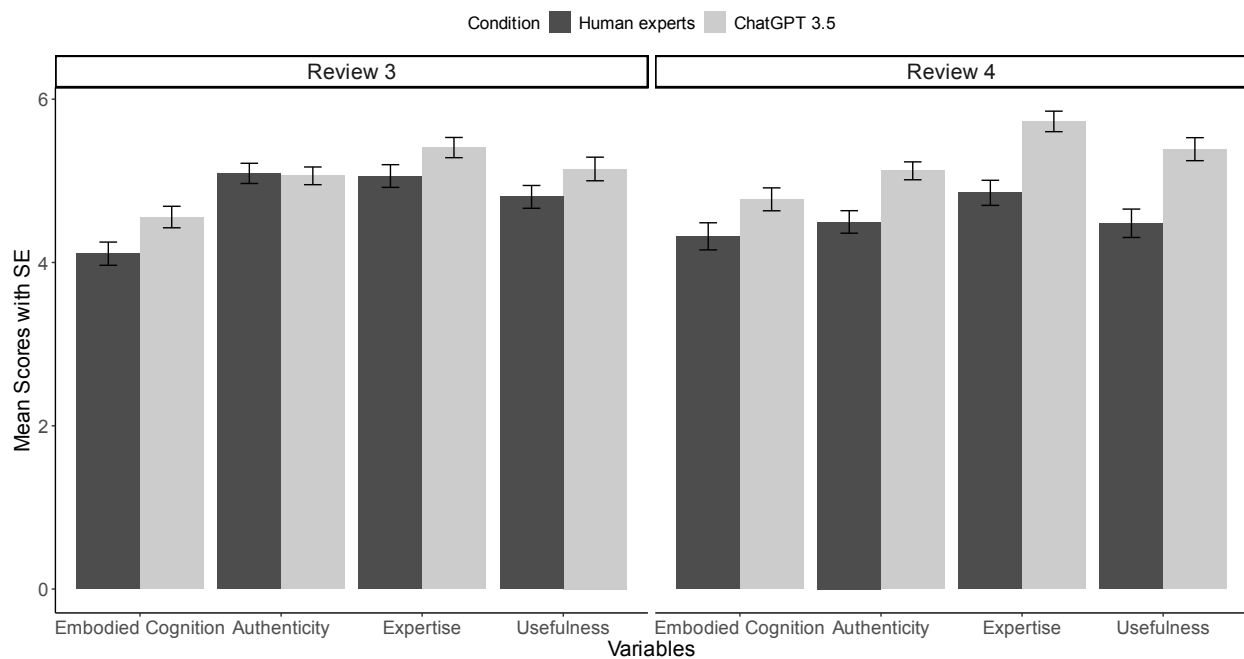
Author Accepted Manuscript

Review 3	5.09	5.06	5.06	5.41	4.80	5.15
	(1.25)	(1.10)	(1.41)	(1.26)	(1.41)	(1.46)
MWU	U = 5046, $p = .63$		U = 5958, $p = .089^\dagger$		U = 5834, $p = .056^\dagger$	
Review 4	4.50	5.12	4.85	5.73	4.48	5.39
	(1.40)	(1.12)	(1.55)	(1.28)	(1.76)	(1.43)
MWU	U = 6668, $p < .001^{***}$		U = 7050, $p < .001^{***}$		U = 6814, $p < .001^{***}$	

Note: Means with standard deviations in parentheses. MWU: Mann-Whitney U tests. $^\dagger p < .10$.

$^{***} p < .001$.

Figure 4. Results of Reviews 3 and 4 in Study 1B



Note: Error bars represent standard errors.

Open-ended question

To understand how consumers perceive the reviews, a further analysis of the open-ended question was conducted. The results showed reading experience and sentence structure played a

Author Accepted Manuscript

role in how participants evaluated the reviews. For instance, one participant mentioned, “*the third review (Review 4) in particular wandered all over the place; the reviewer in my view tried too hard & went off rail into storytelling*”. A comparison of the texts in both conditions uncovered that, unlike human experts who often use succinct and pithy language, ChatGPT reviews included a more reader-friendly structure (nose-palate-finish-summary), making them easier to process. This result highlighted a limitation of relying solely on automatic text analysis tools to evaluate generative AI output: the neglect of context and text structure that are important for readers.

Evidence of hallucination

Generative large language models like ChatGPT suffer from issues of hallucination, which occurs when the model generates seemingly plausible but fabricated content that is not grounded in factual information (Rawte, Sheth and Das 2023). Considering the large number of product reviews in Study 1A, verifying the accuracy of ChatGPT-generated reviews in terms of taste description was challenging. Nevertheless, participants in Study 1B, who were actual alcohol consumers, were able to detect instances of hallucination in ChatGPT reviews. Several participants mentioned that the usage of “smoked bacon” in describing the taste of whisky in Review 3 was peculiar. A word frequency analysis in Study 1A’s reviews revealed that the word “bacon” appeared in 10.36% of the ChatGPT reviews but was scarcely mentioned in human expert reviews (only once as “bacon fat”). The result suggested that content from generative AI had tendencies toward excessive use of specific terms and hallucinations.

Discussion

Author Accepted Manuscript

The results of Study 1B provide the other side of the story from the perspective of consumers. When reading a selection of the reviews from Study 1A, consumers found that certain reviews by human experts displayed lower levels of embodied cognition, expertise, and usefulness, potentially because human experts used more idiosyncratic structures for reviews. As in Study 1A, they noted that ChatGPT reviews contained more positive emotions. Thus, ChatGPT proved to be more effective in expressing positive feelings and more convincing. However, the participants identified the issue of hallucination in content from generative AI, possibly due to its reliance on word-by-word prediction and vulnerability to noisy training data points (Dziri et al. 2022).

Study 2

Study 2 aims to (1) investigate a more sophisticated, multimodal generative AI model (Bing Chat with ChatGPT 4), (2) explore a different use of LLM in marketing communication by benchmarking the outputs of ChatGPT 4 against social media posts from verified accounts, and (3) examine potential downstream consequences (e.g., engagement and purchase intention) of these marketing communications. Prior research has shown that consumers react differently to branded and unbranded content on social media, demonstrating varying levels of persuasion knowledge (Darmawan and Huh 2022) and engaging differently with emotional expressions in these posts (Holiday et al. 2023). Therefore, we test and compare both types of posts in this study.

Data and Procedure

Author Accepted Manuscript

Texts and images from a branded post from GODIVA (verified account) and an unbranded post from a verified influencer account, both featuring chocolate-tasting experiences, were collected and stored separately. Subsequently, two new posts were generated using Bing Chat (<https://www.bing.com/chat>), a publicly free platform integrating ChatGPT 4. Since ChatGPT 4 is a multimodal model, which means the model can process various forms of multimedia such as images, videos, or audio as input (Rahaman et al. 2023), users can also add images in Bing Chat. The text sample for ChatGPT 4 was created by uploading images of the two Instagram posts from verified accounts, accompanied by the following specific prompt (with a concise conversation style). The number of words was specified to match the length of the original verified accounts' posts:

(for branded content)

Please generate texts for this image on an Instagram post. The texts should be approximately 24 words long. No need to include hashtags. Mention GODIVA in the post.

(for unbranded content)

Please generate texts for this image on an Instagram post. The texts should be approximately 49 words long. No need to include hashtags. Do not use texts that already exist on the Internet.

Afterward, we used an Instagram post generator tool (<https://zeoob.com/generate-instagram-post/>) to create two versions of experimental stimuli. Both versions included the same image, with texts either sourced from the original verified accounts or newly generated by ChatGPT 4. We set the same parameters for the number of likes, comments, and the timing of publication in the tool to control for these factors. Two hundred ten US participants (46.86% female, 51.69% male, 1.45% preferred not to say; $M_{age} = 39.63$, $SD_{age} = 11.73$) from the

Author Accepted Manuscript

CloudResearch platform were recruited for an online experiment. After excluding responses that failed the attention check, the final sample size was 207. Participants were instructed to imagine that they were scrolling through Instagram and came across some posts. They were first randomly assigned to one of the two conditions for branded content (human copywriter vs. ChatGPT 4), in which they viewed a post featuring the same image but with different texts (see Web Appendix C for stimuli). After reading the post, subjects rated the level of embodied cognition ($\alpha = .87$; same items as in Study 1B), affect (“How positive do you find the tone/emotions of the Instagram post?” 1 = Not positive at all, 7 = Very positive), and lexical diversity ($\alpha = .80$; same items as in Study 1B) on 7-point Likert scales. Then, they answer questions on how likely they were to engage with this post (e.g., like, comment on it) and to buy the product in the post in the future (1 = Not at all likely, 7 = Very likely). Other variables like authenticity, readability, quality of the content, and brand preference were also measured. In the second part of the study, participants were again randomly assigned to one of the two conditions for unbranded content and read posts from either the influencer account (human copywriter) or generative AI (ChatGPT 4). They responded to the same questions as in the first part, except that the brand preference question was removed. At the end of the study, information on Instagram usage and demographics were collected.

Results

Shapiro-Wilk test results indicated that the distributions of embodied cognition, affect, and lexical diversity for both branded and unbranded content deviated from normality, with all p -values being less than .05. Therefore, the Mann-Whitney U tests were used to compare the group differences as in Studies 1A-B. There was no significant difference in Instagram usage between

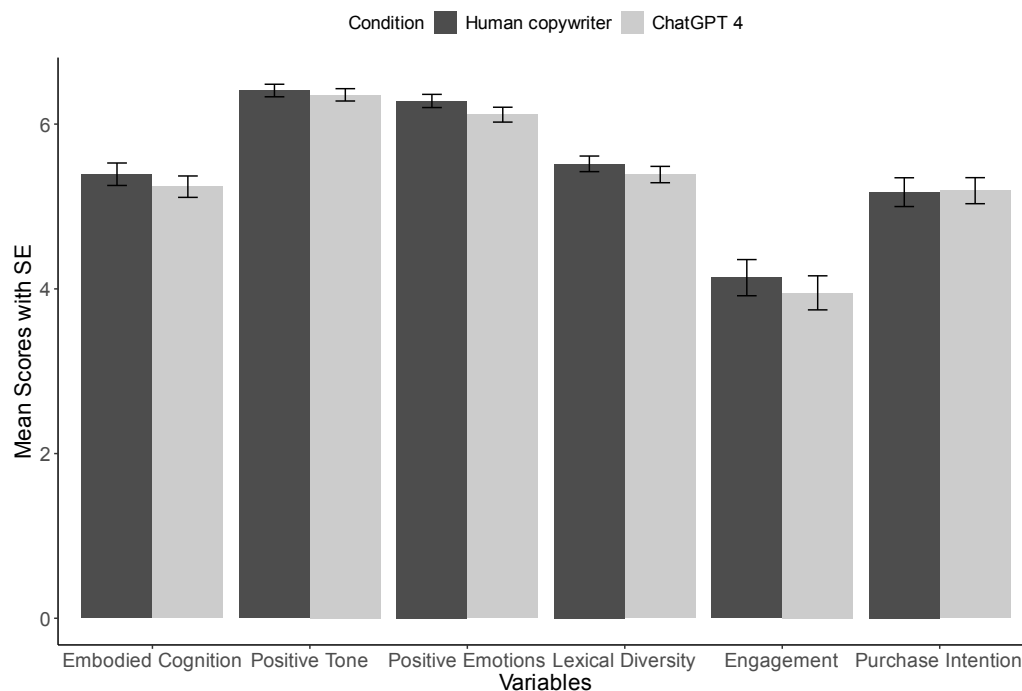
Author Accepted Manuscript

subjects in the human copywriter and ChatGPT 4 conditions (Fisher's exact test: branded content: $p = .56$; unbranded content: $p = .24$).

Branded content

Unlike the findings from Studies 1A-B, there were no significant differences observed in the levels of embodied cognition ($U = 4888.5, p = .28$), positive tone ($U = 5086.5, p = .49$), positive emotions ($U = 4870, p = .23$), or lexical diversity ($U = 4971.5, p = .37$) between participants in the human copywriter and ChatGPT 4 conditions (see Figure 5). Similarly, participants in both conditions did not show significant differences in engagement ($U = 5074, p = .51$), purchase intention ($U = 5210.5, p = .73$), or other variables.

Figure 5. Results of Branded Content in Study 2



Note: Error bars represent standard errors.

Author Accepted Manuscript

The result indicated that ChatGPT 4 could closely mimic the performance of human copywriters in terms of creating social media content and elicit similar levels of cognitive and emotional response from the audience. The lack of significant differences in downstream consequences (i.e., engagement and purchase intention) further suggested that AI-generated content was nearly as effective as content created by human copywriters. This parity in performance across various metrics showed that more sophisticated generative AI could match the capabilities of human copywriters for branded content creation in digital marketing.

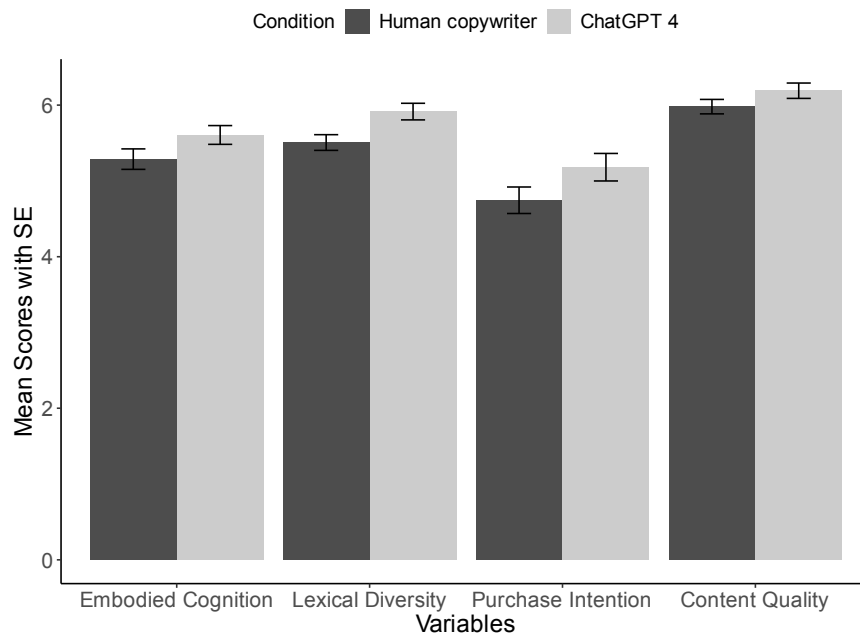
Unbranded content

For unbranded content, there was a marginally significant difference between the two conditions regarding embodied cognition. Compared with the original post from the human copywriter ($M = 5.29$, $SD = 1.36$), the post generated by ChatGPT 4 ($M = 5.61$, $SD = 1.27$) received a slightly higher score in embodied cognition, $U = 6134$, $p = .069$ (see Figure 6). No significant difference was found for positive tone ($U = 5243$, $p = .78$) or positive emotions ($U = 5189.5$, $p = .68$). Interestingly, the ChatGPT 4 post ($M = 5.91$, $SD = 1.12$) had a significantly higher level of lexical diversity than the human copywriter post ($M = 5.51$, $SD = 1.05$), $U = 6708.5$, $p = .0015$ (see Figure 6).

Participants demonstrated an equal likelihood of engagement with the post ($U = 5739.5$, $p = .37$). However, a significant difference was observed in their purchase intention ($U = 6271$, $p = .03$). Subjects in the ChatGPT 4 condition ($M = 5.18$, $SD = 1.86$) had significantly higher purchase intention compared with those in the human copywriter condition ($M = 4.74$, $SD = 1.77$, see Figure 6). The quality of the content was also considered relatively higher for the content produced by ChatGPT 4 ($M = 6.19$, $SD = 1.04$) than by the human copywriter ($M = 5.98$, $SD = 0.97$), $U = 6184$, $p = .04$. No significant difference was found for other variables.

Author Accepted Manuscript

Figure 6. Results of Unbranded Content in Study 2



Note: Error bars represent standard errors.

The results for unbranded content revealed the powerful marketing writing capabilities of generative AI as in Study 1B. ChatGPT 4 was able to generate texts with slightly higher levels of embodied cognition and lexical diversity. Moreover, human judges rated the post with text by ChatGPT 4 as having better quality and expressed a higher purchase intention towards it. The findings implied that ChatGPT 4 might surpass certain Instagram copywriters.

Regression analysis

Regression analysis was conducted to examine the impact of key variables (embodied cognition, affect, and lexical diversity) on engagement and purchase intention, with the key variables as independent variables and engagement/purchase intention as dependent variables.

Both embodied cognition and lexical diversity were significant predictors for engagement ($B_{\text{embodied cognition}} = 0.78, t(202) = 7.21, p < .001$; $B_{\text{lexical diversity}} = 0.40, t(202) = 2.70, p < .01$) and

Author Accepted Manuscript

purchase intention ($B_{\text{embodied cognition}} = 0.68, t(202) = 8.63, p < .001$; $B_{\text{lexical diversity}} = 0.27, t(202) = 2.55, p < .05$) for branded content. For unbranded content, embodied cognition was the only significant predictor for engagement ($B_{\text{embodied cognition}} = 0.93, t(202) = 7.66, p < .001$) and purchase intention ($B_{\text{embodied cognition}} = 0.87, t(202) = 9.58, p < .001$).

Discussion

Study 2 presents intriguing findings by contrasting the writing of human copywriters from verified social media accounts with that of a more advanced AI model, ChatGPT 4, as perceived by consumers. For branded content, participants were unable to discern any significant differences between content created by the brand and that generated by the AI model. The results were even more striking for unbranded content. ChatGPT 4's creation was not only comparable in terms of intended engagement and purchase intention but was also perceived as superior in embodied cognition, lexical diversity, and content quality. The absence of significant differences in the branded content segment could suggest that ChatGPT's capabilities are on par with those of professional copywriters, as brands typically employ professional writers for their content. Conversely, for non-branded social media accounts, where copywriting is not always done by professionals, the distinctions in content quality become more pronounced. It should be noted, though, that the publication dates of the original Instagram posts are different for branded and unbranded content (2023 for branded content and 2021 for unbranded content). Since the branded content was published after ChatGPT's launch, there is a possibility that the brand started using ChatGPT for its social media posts.

General Discussion

Author Accepted Manuscript

Can generative AI describe a sensory consumption experience, such as tasting whiskey or chocolate, as humans do? The current research addresses this question by comparing content from humans and two consecutive generations of generative AI (ChatGPT 3.5 and 4) across three important capabilities to describe sensory experiences: (1) embodying cognition from (or as if from) physical experiences, (2) conveying affect through emotions and tones, and (3) applying a diverse and nuanced vocabulary for sensory experiences. Through a multi-method approach involving both automatic text analysis and human evaluations, the studies provide insights at both macro and micro levels. While automatic textual analysis across a broad corpus revealed that human experts demonstrated higher levels of embodied cognition and lexical diversity, they were less expressive in affect. Conversely, when given only a sample of the content from either human experts or generative AI, human raters were unable to detect differences except for positive emotions. Generative AI output was perceived as matching or even exceeding human writers' capabilities in creating engaging narratives and persuading consumers. In two studies, generative AI exhibits more positive emotions/tones than human experts.

As far as we know, this is the first research to investigate the application of generative AI in creating experiential narratives in marketing and its downstream impact, including consumer responses and marketing performance outcomes (e.g., intended purchase and engagement). Our findings provide insights into how generative AI transforms marketing communication by studying the experiential narratives created by generative AI, from content characteristics to consumer reactions, and comparing them with human-generated content. This work lays the groundwork for future studies in the intersection of artificial intelligence and marketing practices. Based on objective automatic text analysis of a large corpus, human-generated

Author Accepted Manuscript

experiential narratives exhibit superior embodied cognition and vocabulary diversity. Contrary to prior research, this study demonstrates that generative AI can convey positive emotional experiences to consumers. In essence, generative AI demonstrates a form of artificial empathy, marking a significant step forward for the next generation of AI in marketing communications (Liu-Thompkins et al. 2022). At last, the current paper sheds light on how different aspects of experiential narratives impact consumer intended engagement. It suggests that the level of embodied cognition and the diversity of sensory language have a more substantial effect on positive consumer attitudes and engagement than positive affect.

Considerations for Public Policy

Our findings highlight some implications of the growing use of generative AI by firms to create a massive amount of marketing content at a lower cost. This paper demonstrates the capability of large language models to mimic and even outperform professional copywriters in crafting compelling yet potentially misleading narratives targeted at consumers. It underscores the potential role of these models in spreading misinformation about products and services either by including accidental hallucinations (like in Study 1B for product reviews) or being intentionally used by brands and influencers. Legitimized influencer accounts have also been identified as significant sources for disseminating misinformation on social media platforms (Di Domenico, Nunan and Pitardi 2022). Given the considerable influence of online consumer reviews and influencer content on consumer decisions, the use of generative AI by brands to create a large number of deceptive reviews with minimal effort and costs raises significant ethical concerns. Unable to discern that a review or post has been generated by ChatGPT rather than a human, consumers may base their decisions on fake content.

Author Accepted Manuscript

To mitigate these risks, policymakers must deliberate on preemptive measures, including mandatory disclosure of AI involvement in content creation. The recent EU Artificial Intelligence Act (<https://artificialintelligenceact.eu/>), the first comprehensive regulation on AI, provides a framework for thinking about risks related to AI applications. Most of the text addresses prohibited and high-risk AI systems. Prohibited AI systems refer to social scoring systems and manipulative AI, while high-risk AI systems are those profiling individuals by the automated processing of personal data. The use of generative AI for marketing communication seems to belong to the limited-risk category for which the EU Act is only calling for more transparency: “Deployers of an AI system that generates or manipulates text which is published with the purpose of informing the public on matters of public interest shall disclose that the text has been artificially generated or manipulated.” The introduction of warning labels for AI-generated content in marketing communications is thus recommended to safeguard transparency and consumer rights to make informed decisions. The need for transparency and regulation of generative AI is especially crucial to protect vulnerable consumer segments (Weber et al. 2021). The labels can include several key aspects such as AI content identification, purpose of use, and disclaimer on accuracy. Such labels that indicate the potential risks of inaccurate facts and hallucinations from AI could serve as an essential tool in combating the spread of misinformation and disinformation (Mende et al. 2024).

Another way to curtail the ethical issues related to the use of generative AI in marketing communication is to ensure some human oversight over the process. The EU AI Act sees it as an alternative to the disclosure obligation: “This obligation shall not apply where [...] the AI-generated content has undergone a process of human review or editorial control and where a natural or legal person holds editorial responsibility for the publication of the content.”

Author Accepted Manuscript

Particularly in product descriptions, the 100% reliance on generative AI without human oversight should be prohibited to ensure content accuracy. Human oversight could encompass specific content creation guidelines or approval workflows to safeguard content quality and veracity.

Yet, the above recommendations rely on the goodwill of companies to implement them. Is it possible to enforce these guidelines about transparency if, as our Study 2 indicates, it is barely possible for a human judge to detect generative AI intervention? The automatic analysis of a large corpus of generated content and its systematic comparison with human-generated content may provide valuable insights into developing AI-detection tools. Using this approach on a small scale in Study 1A, the automated textual analysis revealed several distinctive cues in the experiential narratives generated by ChatGPT: an over-reliance on positive affect, less embodied cognition, and a less varied vocabulary. These differences (and others) may present an opportunity for regulatory bodies to develop algorithms for detecting AI-generated content (see Liang et al., 2023, for an example of such a detector). However, keeping up with the pace of LLM development and progress to emulate human language will represent a huge challenge. The insights gained from these algorithms could also be included in campaigns to educate consumers on how to identify AI-generated content, thereby further minimizing potential risks. Consumer education represents a strategic intervention to elucidate the nature and application of generative AI across various sectors. By leveraging disclosures from companies, regulators can illustrate the role of generative AI in content creation and enhance public understanding. Given that consumers become vigilant only when they are suspicious of the environment (Banker and Khetani 2019), they must be informed that, despite the apparent authenticity of AI-generated content, it does not correspond to a real human experience but is instead derived from statistical

Author Accepted Manuscript

analyses of extensive training data collected from the Internet and various media sources.

Furthermore, regular educational campaigns that increase general literacy on AI technology should be systematically implemented. These campaigns should aim to demystify algorithmic operations, explain the potential for inherent biases stemming from human prejudices in training datasets, and the “black box” nature of these systems, which operates on a “garbage in, garbage out” principle. This should help consumers understand that generative AI outputs can sometimes become unreliable. Enhancing such literacy on AI technology will equip consumers to navigate better the increasingly AI-driven landscape of information and content creation, thereby facilitating more informed and discerning decisions.

Future Research

The present research has several limitations. Some of these limitations can inspire further avenues for research for which we offer proposition in Web Appendix D (see Table W3). First, the scope of this research was limited to two products within the food and beverage category. This type of content sets the bar especially high for generative AI to compare with human-generated content because of its distinctive dimensions (affect, lexical diversity, and embodied cognition). Future studies should investigate a broader array of product categories and examine different narratives by generative AI. Our findings should extend to other types of experiences that involve hedonic dimensions, such as restaurants, hotels, entertainment venues, and cultural products. We expect that the comparison between human and AI content will reveal even fewer differences for utilitarian products, such as home appliances or telecommunication services. These usage experiences typically involve more factual descriptions and evaluations (i.e., less

Author Accepted Manuscript

affect), fewer bodily sensations (i.e., embodied cognition), and likely less variation from one consumer to another (i.e., lower lexical diversity).

Second, there exist discrepancies between the results obtained through automatic text analysis and those derived from human rater evaluations. This is understandable, as automated text analysis provides an objective and quantitative assessment of the texts, which might not align with the subjective perceptions of a human audience. The ratings of AI-generated content by humans require large samples of participants, which is time-consuming and costly. Future research could address this methodological challenge by developing refined text analysis tools that replicate human evaluation. This might involve the use of generative AI as well. Recent studies on the use of generative AI in marketing research suggest promising possibilities for large language models (LLMs) to serve as substitutes for human participants in market research. (Goli and Singh 2024; Li et al. 2024).

Third, the content generation process of our studies depends on the specific prompts we designed and the generative AI models used. Alternative prompts could yield substantially different outputs, potentially altering the results of our investigations. Future research may benefit from a more granular approach to control the outputs of generative AI systems, such as ChatGPT. The prompts could be more explicit about the desired perceptions and effects on the reader to customize the LLM outputs. For instance, future studies could instruct ChatGPT to produce text with various levels of embodied cognition, affect, and lexical diversity. While our study focused on ChatGPT 3.5 and ChatGPT 4, exploring other state-of-the-art models like Google's Gemini and Anthropic's Claude could provide a broader perspective on the progression of AI technologies. Such comparisons would shed light on specific strengths and weaknesses

Author Accepted Manuscript

across models, offering deeper insights into their potential applications in creating diverse types of content and enhancing consumer engagement.

Fourth, this research focused exclusively on textual content. Further investigation could explore the capacity of generative AI tools to create other types of content (audio, visual) and compare them with similar human-generated content. The most recent model, ChatGPT 4, is already multimodal and can process texts, images, sounds, and videos (Rahaman et al. 2023). For example, in Study 2, ChatGPT was instructed to generate text to accompany a picture for an Instagram post. We could envision asking the most recent LLMs to generate pictures, audio, or videos. Applications like DALL-E, which create images from text prompts, demonstrate the potential for such advancements. Virtual influencers represent a marketing application context for video-generated content. With the rise of influencer marketing, companies are increasingly deploying virtual influencers across social marketing platforms (Kim and Wang 2023), despite research indicating consumers' skepticism towards virtual influencers' proximal sensory (e.g., olfactory) capabilities (Zhou, Yan and Jiang 2023). Leveraging ostensibly legitimized content from generative AI, virtual influencers can craft compelling narratives that potentially alter consumer intentions and behaviors.

Fifth, another future research direction is to examine the impact of disclosing content origins (AI versus humans) on consumer perceptions. In the present studies, participants were not informed of the potential AI authorship because, consistent with our research question, we did not want their evaluations to be biased by the source cue. A different research question would be to investigate how the disclosure of generative AI involvement, consistent with the EU AI Act mentioned earlier, affects consumer perceptions of the content. Previous research indicates that AI authorship does not significantly alter text perception, though AI authors are

Author Accepted Manuscript

regarded as less anthropomorphic and intelligent than human authors (Lermann Henestrosa, Greving and Kimmerle 2023). Investigations on whether revealing AI authorship before and after reading the texts may shed light on the complex interactions between the source of content, text perception, behavioral intentions, and author perception.

Conclusion

In this paper, we explored the capabilities of generative AI in creating experiential narratives within the realm of marketing communications and compare its performance with that of human-created content on embodied cognition, affect, and lexical diversity. Our analysis revealed nuanced insights into strengths and weaknesses in generative AI's ability to emulate human-like experiential narratives in marketing communications. The findings demonstrate that generative AI can match or even exceed human writers across three key dimensions and generate compelling narratives to drive purchase intention. Nonetheless, a notable risk associated with AI-generated content is the possibility of producing and disseminating inaccurate information that is not based on any real product experience. This paper not only highlights the growing effectiveness of generative AI in replicating human emotional and sensory experiences but also emphasizes the need for transparency and ethical deployment. Looking forward, the paper suggests avenues for future research to ensure responsible AI integration in marketing. This exploration serves as a foundation for both marketing practitioners and policymakers, providing valuable guidance on the responsible harnessing of AI's immense potential in shaping future marketing landscapes.

Author Accepted Manuscript

References

- AlAfnan, Mohammad Awad and Siti Fatimah MohdZuki (2023), “Do Artificial Intelligence Chatbots Have a Writing Style? An Investigation into the Stylistic Features of ChatGPT-4,” *Journal of Artificial Intelligence and Technology*, 3(3), 85–94.
- Anderson, Michael L., Michael J. Richardson, and Anthony Chemero (2012), “Eroding the Boundaries of Cognition: Implications of Embodiment1,” *Topics in Cognitive Science*, 4(4), 717–30.
- Banker, Sachin and Salil Khetani (2019), “Algorithm Overdependence: How the Use of Algorithmic Recommendation Systems Can Increase Risks to Consumer Well-Being,” *Journal of Public Policy & Marketing*, 38(4), 500–515.
- Barsalou, Lawrence W. (2008), “Grounded Cognition,” *Annual Review of Psychology*, 59(Volume 59, 2008), 617–45.
- Berger, Jonah and Katherine L Milkman (2012), “What Makes Online Content Viral?,” *Journal of Marketing Research*, 49(2), 192–205.
- Beverland, Michael B. and Francis J. Farrelly (2010), “The Quest for Authenticity in Consumption: Consumers’ Purposive Choice of Authentic Cues to Shape Experienced Outcomes,” *Journal of Consumer Research*, 36(5), 838–56.
- Boyd, Ryan L, Ashwini Ashokkumar, Sarah Seraj, and James W Pennebaker (2022), “The Development and Psychometric Properties of LIWC-22,” *Austin, TX: University of Texas at Austin*, 1–47.
- Brakus, J. Joško, Bernd H. Schmitt, and Lia Zarantonello (2009), “Brand Experience: What Is It? How Is It Measured? Does It Affect Loyalty?,” *Journal of Marketing*, 73(3), 52–68.
- Cao, Yihan, Siyu Li, Yixin Liu, Zhiling Yan, Yutong Dai, Philip S. Yu, and Lichao Sun (2023), “A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT,” <http://arxiv.org/abs/2303.04226>.
- Cardona, Juan F., Lucila Kargieman, Vladimiro Sinay, Oscar Gershanik, Carlos Gelormini, Lucia Amoroso, María Roca, David Pineda, Natalia Trujillo, Maëva Michon, Adolfo M. García, Daniela Szenkman, Tristán Bekinschtein, Facundo Manes, and Agustín Ibáñez (2014), “How Embodied Is Action Language? Neurological Evidence from Motor Diseases,” *Cognition*, 131(2), 311–22.
- Cascio Rizzo, Giovanni Luca, Jonah Berger, Matteo De Angelis, and Rumen Pozharliev (2023), “How Sensory Language Shapes Influencer’s Impact,” *Journal of Consumer Research*, 50(4), 810–25.

Author Accepted Manuscript

- Darmawan, Ida and Jisu Huh (2022), “The Effects of Message Type and Sponsorship Disclosure in Influencer Marketing of Prescription Drugs,” *Journal of Global Marketing*, 35(1), 21–36.
- Davenport, Thomas, Abhijit Guha, Dhruv Grewal, and Timna Bressgott (2020), “How Artificial Intelligence Will Change the Future of Marketing,” *Journal of the Academy of Marketing Science*, 48(1), 24–42.
- De Freitas, Julian, Ahmet Kaan Uğuralp, Zeliha Oğuz-Uğuralp, and Stefano Puntoni (2023), “Chatbots and Mental Health: Insights into the Safety of Generative AI,” *Journal of Consumer Psychology*.
- Di Domenico, Giandomenico, Daniel Nunan, and Valentina Pitardi (2022), “Marketplaces of Misinformation: A Study of How Vaccine Misinformation Is Legitimized on Social Media,” *Journal of Public Policy & Marketing*, 41(4), 319–35.
- Digutsch, Jan and Michal Kosinski (2023), “Overlap in Meaning Is a Stronger Predictor of Semantic Activation in GPT-3 than in Humans,” *Scientific Reports*, 13(1), 1–7.
- Downes, William (2000), “The Language of Felt Experience: Emotional, Evaluative and Intuitive,” *Language and Literature*, 9(2), 99–121.
- Dziri, Nouha, Sivan Milton, Mo Yu, Osmar Zaiane, and Siva Reddy (2022), “On the Origin of Hallucinations in Conversational Models: Is It the Datasets or the Models?,” <http://arxiv.org/abs/2204.07931>.
- Eckler, Petya and Paul Bolls (2011), “Spreading the Virus,” *Journal of Interactive Advertising*, 11(2), 1–11.
- Elder, Ryan S. and Aradhna Krishna (2011), “The ‘Visual Depiction Effect’ in Advertising: Facilitating Embodied Mental Simulation through Product Orientation,” *Journal of Consumer Research*, 38(6), 988–1003.
- Escalas, Jennifer Edson (2004), “Narrative Processing: Building Consumer Connections to Brands,” *Journal of Consumer Psychology*, 14(1), 168–80.
- Eulerich, Marc, Aida Sanatizadeh, Hamid Vakilzadeh, and David A. Wood (2023), “Is It All Hype? ChatGPT’s Performance and Disruptive Potential in the Accounting and Auditing Industries,” <https://papers.ssrn.com/abstract=4452175>.
- Foglia, Lucia and Robert A. Wilson (2013), “Embodied Cognition,” *WIREs Cognitive Science*, 4(3), 319–25.
- Goli, Ali and Amandeep Singh (2024), “Frontiers: Can Large Language Models Capture Human Preferences?,” *Marketing Science*.

Author Accepted Manuscript

- Herbold, Steffen, Annette Hautli-Janisz, Ute Heuer, Zlata Kikteva, and Alexander Trautsch (2023), "A Large-Scale Comparison of Human-Written versus ChatGPT-Generated Essays," *Scientific Reports*, 13(1), 18617.
- Holiday, Steven, Jameson L. Hayes, Haseon Park, Yuanwei Lyu, and Yang Zhou (2023), "A Multimodal Emotion Perspective on Social Media Influencer Marketing: The Effectiveness of Influencer Emotions, Network Size, and Branding on Consumer Brand Engagement Using Facial Expression and Linguistic Analysis," *Journal of Interactive Marketing*, 58(4), 414–39.
- Huang, Ming-Hui and Roland T. Rust (2018), "Artificial Intelligence in Service," *Journal of Service Research*, 21(2), 155–72.
- Jarvis, Scott (2013), "Capturing the Diversity in Lexical Diversity," *Language Learning*, 63(s1), 87–106.
- Johnson, Wendell (1944), "Studies in Language Behavior: A Program of Research," *Psychological Monographs*, 56(2), 1–15.
- Katz, Daniel Martin, Michael James Bommarito, Shang Gao, and Pablo Arredondo (2023), "GPT-4 Passes the Bar Exam," <https://papers.ssrn.com/abstract=4389233>.
- Kim, Donggyu and Zituo Wang (2023), "The Ethics of Virtuality: Navigating the Complexities of Human-like Virtual Influencers in the Social Media Marketing Realm," *Frontiers in Communication*, 8, <https://www.frontiersin.org/articles/10.3389/fcomm.2023.1205610>.
- Krishna, Aradhna and Norbert Schwarz (2014), "Sensory Marketing, Embodiment, and Grounded Cognition: A Review and Introduction," *Journal of Consumer Psychology*, 24(2), 159–68.
- Kumar, Ashish, Ram Bezawada, Rishika Rishika, Ramkumar Janakiraman, and P.K. Kannan (2016), "From Social to Sale: The Effects of Firm-Generated Content in Social Media on Customer Behavior," *Journal of Marketing*, 80(1), 7–25.
- Kwon, Mina and Rashmi Adaval (2018), "Going against the Flow: The Effects of Dynamic Sensorimotor Experiences on Consumer Choice," *Journal of Consumer Research*, 44(6), 1358–78.
- Lermann Henestrosa, Angelica, Hannah Greving, and Joachim Kimmerle (2023), "Automated Journalism: The Effects of AI Authorship and Evaluative Information on the Perception of a Science Journalism Article," *Computers in Human Behavior*, 138, 107445.
- Li, Peiyao, Noah Castelo, Zsolt Katona, and Miklos Sarvary (2024), "Frontiers: Determining the Validity of Large Language Models for Automated Perceptual Analysis," *Marketing Science*, 43(2), 254–66.
- Liang, Gongbo, Jesus Guerrero, Fengbo Zheng, and Izzat Alsmadi (2023), "Enhancing Neural Text Detector Robustness with μ Attacking and RR-Training," *Electronics*, 12(8), 1948.

Author Accepted Manuscript

- Liang, Paul Pu, Chiyu Wu, Louis-Philippe Morency, and Ruslan Salakhutdinov (2021), "Towards Understanding and Mitigating Social Biases in Language Models," in *Proceedings of the 38th International Conference on Machine Learning*, PMLR, 6565–76, <https://proceedings.mlr.press/v139/liang21a.html>.
- Libai, Barak, Yakov Bart, Sonja Gensler, Charles F. Hofacker, Andreas Kaplan, Kim Kötterheinrich, and Eike Benjamin Kroll (2020), "Brave New World? On AI and the Management of Customer Relationships," *Journal of Interactive Marketing*, 51(1), 44–56.
- Liu, Angela Xia, Ying Xie, and Jurui Zhang (2019), "It's Not Just What You Say, But How You Say It: The Effect of Language Style Matching on Perceived Quality of Consumer Reviews," *Journal of Interactive Marketing*, 46(1), 70–86.
- Liu, Xiaohui, Na Jiang, Haiping Zhao, Eric Lim, and Chee-Wee Tan (2023), "Describing Your Podcasts: Disentangling the Effects of Linguistic Features in Bolstering Online Learners' Engagement with Podcasters," *PACIS 2023 Proceedings*, <https://aisel.aisnet.org/pacis2023/194>.
- Liu-Thompkins, Yuping, Shintaro Okazaki, and Hairong Li (2022), "Artificial Empathy in Marketing Interactions: Bridging the Human-AI Gap in Affective and Social Customer Experience," *Journal of the Academy of Marketing Science*, 50(6), 1198–1218.
- Lovett, Mitchell J., Renana Peres, and Ron Shachar (2013), "On Brands and Word of Mouth," *Journal of Marketing Research*, 50(4), 427–44.
- Lynch, Christopher J., Erik J. Jensen, Virginia Zamponi, Kevin O'Brien, Erika Frydenlund, and Ross Gore (2023), "A Structured Narrative Prompt for Prompting Narratives from Large Language Models: Sentiment Assessment of ChatGPT-Generated Narratives and Real Tweets," *Future Internet*, 15(12), 375.
- Ma, Liye and Baohong Sun (2020), "Machine Learning and AI in Marketing – Connecting Computing Power to Human Insights," *International Journal of Research in Marketing*, 37(3), 481–504.
- Mahowald, Kyle, Anna A. Ivanova, Idan A. Blank, Nancy Kanwisher, Joshua B. Tenenbaum, and Evelina Fedorenko (2023), "Dissociating Language and Thought in Large Language Models," <http://arxiv.org/abs/2301.06627>.
- Maleki, Negar, Balaji Padmanabhan, and Kaushik Dutta (2024), "AI Hallucinations: A Misnomer Worth Clarifying," <http://arxiv.org/abs/2401.06796>.
- McCarthy, Philip M. and Scott Jarvis (2010), "MTLD, Vocd-D, and HD-D: A Validation Study of Sophisticated Approaches to Lexical Diversity Assessment," *Behavior Research Methods*, 42(2), 381–92.

Author Accepted Manuscript

- Mende, Martin, Valentina O. Ubal, Marina Cozac, Beth Vallen, and Christopher Berry (2024), "Fighting Infodemics: Labels as Antidotes to Mis- and Disinformation?!", *Journal of Public Policy & Marketing*, 43(1), 31–52.
- Mishra, Chinmaya, Rinus Verdonshot, Peter Hagoort, and Gabriel Skantze (2023), "Real-Time Emotion Generation in Human-Robot Dialogue Using Large Language Models," *Frontiers in Robotics and AI*, 10, 1271610.
- Mora, Pierre and Jean Moscarola (2010), "Representations of the Emotions Associated with a Wine Purchasing or Consumption Experience," *International Journal of Consumer Studies*, 34(6), 674–83.
- Nunes, Joseph C., Andrea Ordanini, and Francesca Valsesia (2015), "The Power of Repetition: Repetitive Lyrics in a Song Increase Processing Fluency and Drive Market Success," *Journal of Consumer Psychology*, 25(2), 187–99.
- OpenAI (2022), "Introducing ChatGPT," <https://openai.com/blog/chatgpt>.
- Poels, Karolien and Siegfried Dewitte (2019), "The Role of Emotions in Advertising: A Call to Action," *Journal of Advertising*, 48(1), 81–90.
- Press, Gil (2023), "Making ChatGPT A Highly-Targeted Marketing Tool, Anyword Predicts Its Performance," *Forbes*, <https://www.forbes.com/sites/gilpress/2023/07/18/making-chatgpt-a-highly-targeted-marketing-tool-anyword-predicts-its-performance/>.
- Rahaman, Md Saidur, MM Tahmid Ahsan, Nishath Anjum, Harold Jan R Terano, and Md Mizanur Rahman (2023), "From ChatGPT-3 to GPT-4: A Significant Advancement in Ai-Driven NLP Tools," *Journal of Engineering and Emerging Technologies*, 2(1), 1–11.
- Rawte, Vipula, Amit Sheth, and Amitava Das (2023), "A Survey of Hallucination in Large Foundation Models," <http://arxiv.org/abs/2309.05922>.
- Ray, Partha Pratim (2023), "ChatGPT: A Comprehensive Review on Background, Applications, Key Challenges, Bias, Ethics, Limitations and Future Scope," *Internet of Things and Cyber-Physical Systems*, 3, 121–54.
- Rocklage, Matthew D. and Russell H. Fazio (2020), "The Enhancing Versus Backfiring Effects of Positive Emotion in Consumer Reviews," *Journal of Marketing Research*, 57(2), 332–52.
- Rocklage, Matthew D, Derek D Rucker, and Loran F Nordgren (2021), "Emotionally Numb: Expertise Dulls Consumer Experience," *Journal of Consumer Research*, 48(3), 355–73.
- Russell, James A. (2003), "Core Affect and the Psychological Construction of Emotion," *Psychological Review*, 110(1), 145–72.
- Schermerhorn, Vaughn (2023), "How Amazon Continues to Improve the Customer Reviews Experience with Generative AI," *About Amazon*,

Author Accepted Manuscript

<https://www.aboutamazon.com/news/amazon-ai/amazon-improves-customer-reviews-with-generative-ai>.

- Schmitt, Bernd (1999), “Experiential Marketing,” *Journal of Marketing Management*, 15(1–3), 53–67.
- Srivastava, Aarohi, Abhinav Rastogi, Abhishek Rao, Abu Awal Md Shoeb, Abubakar Abid, Adam Fisch, Adam R Brown, Adam Santoro, Aditya Gupta, Adrià Garriga-Alonso, and others (2022), “Beyond the Imitation Game: Quantifying and Extrapolating the Capabilities of Language Models,” *arXiv preprint arXiv:2206.04615*.
- Tak, Ala N. and Jonathan Gratch (2023), “Is GPT a Computational Model of Emotion? Detailed Analysis,” <http://arxiv.org/abs/2307.13779>.
- Takagi, Soshi, Takashi Watari, Ayano Erabi, and Kota Sakaguchi (2023), “Performance of GPT-3.5 and GPT-4 on the Japanese Medical Licensing Examination: Comparison Study,” *JMIR Medical Education*, 9(1), e48002.
- Tseng, Rayden, Suzan Verberne, and Peter van der Putten (2023), “ChatGPT as a Commenter to the News: Can LLMs Generate Human-Like Opinions?,” in *Disinformation in Open Online Media*, ed. Davide Ceolin, Tommaso Caselli, and Marina Tulin, Cham: Springer Nature Switzerland, 160–74.
- Ullman, Tomer (2023), “Large Language Models Fail on Trivial Alterations to Theory-of-Mind Tasks,” <http://arxiv.org/abs/2302.08399>.
- Verma, Pranshu and Gerrit De Vynck (2023), “ChatGPT Took Their Jobs. Now They Walk Dogs and Fix Air Conditioners.,” *Washington Post*, June 5, <https://www.washingtonpost.com/technology/2023/06/02/ai-taking-jobs/>.
- Webb, Taylor, Keith J. Holyoak, and Hongjing Lu (2023), “Emergent Analogical Reasoning in Large Language Models,” *Nature Human Behaviour*, 7(9), 1526–41.
- Weber, T.J., Chris Hydock, William Ding, Meryl Gardner, Pradeep Jacob, Naomi Mandel, David E. Sprott, and Eric Van Steenburg (2021), “Political Polarization: Challenges, Opportunities, and Hope for Consumer Welfare, Marketers, and Public Policy,” *Journal of Public Policy & Marketing*, 40(2), 184–205.
- Wilson, Margaret (2002), “Six Views of Embodied Cognition,” *Psychonomic Bulletin & Review*, 9(4), 625–36.
- Xu, Qihui, Yingying Peng, Samuel A. Nastase, Martin Chodorow, Minghua Wu, and Ping Li (2023), “Does Conceptual Representation Require Embodiment? Insights From Large Language Models,” <http://arxiv.org/abs/2305.19103>.
- Zhou, Xinyue, Xiao Yan, and Yuwei Jiang (2023), “Making Sense? The Sensory-Specific Nature of Virtual Influencer Effectiveness,” *Journal of Marketing*, 84–106.

Author Accepted Manuscript

Zhu, Feng and Xiaoquan (Michael) Zhang (2010), “Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics,” *Journal of Marketing*, 74(2), 133–48.

Zhuang, Yan, Qi Liu, Yuting Ning, Weizhe Huang, Rui Lv, Zhenya Huang, Guanhao Zhao, Zheng Zhang, Qingyang Mao, Shijin Wang, and Enhong Chen (2023), “Efficiently Measuring the Cognitive Ability of LLMs: An Adaptive Testing Perspective,” <http://arxiv.org/abs/2306.10512>.

Peer Review Version

Author Accepted Manuscript

Web Appendix: Experiential Narratives in Marketing: A Comparison of Generative AI and Human Content

Table of Contents

<i>Web Appendix A: Summary Statistics of Review Samples in Study 1A</i>	2
<i>Web Appendix B: Review Sample in Study 1B</i>	3
<i>Web Appendix C: Stimuli in Study 2 (Branded and Unbranded Content)</i>	6
<i>Web Appendix D: Propositions for Further Research</i>	7

Yingting Wen
EMLYON Business School
144 avenue Jean Jaurès, 69007 Lyon, France
wen@em-lyon.com

Sandra Laporte
Toulouse School of Management
University Toulouse Capitole
2 rue du Doyen Gabriel Marty, 31042 Toulouse, France
sandra.laporte@tsm-education.fr

These materials have been supplied by the authors to aid in the understanding of their paper. The AMA is sharing these materials at the request of the authors.

Author Accepted Manuscript

Web Appendix A: Summary Statistics of Review Samples in Study 1A

Table W1. Summary Statistics of Review Samples in Study 1A

Variables	Minimum	25th percentile	Median	75th percentile	Maximum
Product price	13	70	117	200	26650
Review point	73	84	87	89	96
Word count					
<i>Human Experts</i>	24	63	72	78	169
<i>ChatGPT 3.5</i>	53	64	68	77	118
Embodied cognition					
<i>Human Experts</i>	0	12.53	15.90	18.05	29.17
<i>ChatGPT 3.5</i>	1.45	6.59	9.23	12.98	22.37
Affect					
<i>Human Experts</i>	0	2.78	4.62	6.58	14.29
<i>ChatGPT 3.5</i>	0	4.48	6.08	7.46	14.29
Positive tones					
<i>Human Experts</i>	0	2.54	3.90	6.25	13.46
<i>ChatGPT 3.5</i>	0	3.49	5.52	7.06	14.29
Positive emotions					
<i>Human Experts</i>	0	0	0	1.27	6.12
<i>ChatGPT 3.5</i>	0	0	1.34	1.77	5.19

Note: Review point refers to the quantitative score assigned to the reviewed product on a 100-point scale.

Web Appendix B: Review Sample in Study 1B**Table W2.** Review Sample in Study 1B

Reviews	Human Experts	ChatGPT 3.5
Review 1	<p>The ruby/antique amber color suggests a first-fill sherry cask; the nose and palate confirm it. This whisky is a mouthful: fruitcake, roasted chestnuts, burnt raisin, candied orange, molasses, leather, tobacco, dark chocolate, wax, and a hint of clove. An incredible amount of sherry, but it's not cloying. The flavors are clean, complex, and well-balanced. (LIWC scores: embodied cognition = 10.9; positive emotion = 0)</p>	<p>Murray McDavid's Mission Glenglassaugh 1986 Vintage 20 year old, at 55.3%, offers a nose with tropical fruit notes, vanilla, and a hint of oak. The palate is a delightful blend of flavors—ripe mango, coconut, and a touch of white chocolate create a complex and well-rounded profile. The finish is medium-length, with a final note of apricots. A rare and flavorful Highland single malt. (LIWC scores: embodied cognition = 4.69; positive emotion = 1.56)</p>
Review 2	<p>Both bourbon and sherry casks were used for aging. Stewed fruits, butterscotch, and soft spices on the nose. Silky palate delivery, with honey, dark chocolate, light sherry, and cloves. Lingering finish, with citrus and allspice. (LIWC scores: embodied</p>	<p>Aberlour's 16 year old is a Speyside classic. At 40%, it offers a nose with hints of sherry sweetness, dried fruits, and a touch of oak. The palate is a delightful mix of flavors—sultanas, Christmas cake, and a gentle spice create a well-rounded and flavorful</p>

Author Accepted Manuscript

cognition = 14.29; positive emotion = 0)

profile. The finish is medium-length, leaving a hint of cinnamon and dark chocolate. A well-crafted and approachable single malt. (LIWC scores: embodied cognition = 6.15; positive emotion = 1.54)

Review 3 Aged in a sherry cask, and the sherry is clean and bright. I'm also picking up a good dose of warming peat smoke, along with molasses, and a gentle maltiness throughout. A peppering of subtle brine and a hint of seaweed peak through occasionally, reminding one of its Islay roots. (LIWC scores: embodied cognition = 16; positive emotion = 2)

Bruichladdich's Infinity is an Islay delight. At 55.5%, it presents a nose filled with peat smoke, maritime notes, and hints of brine. The palate is a bold and harmonious blend of flavors—smoked bacon, seaweed, and a touch of citrus create a rich and layered profile. The finish is long and smoky, celebrating Islay's peaty character. A well-rounded and flavorful whisky. (LIWC scores: embodied cognition = 6.56; positive emotion = 3.28)

Review 4 Peat has grown up to become the best version of himself that he can be. The mellow, sweet smoke makes all your cares drift away. It's like walking on an





Big Peat 25-year-old is a robust and smoky delight. Its golden hue hints at the complexity within. The nose offers a bold mix of peat smoke, brine, and a

Author Accepted Manuscript

Islay beach through seaweed stranded touch of medicinal notes. On the palate,
at the high-tide mark. Lemon lozenges, it unfolds with layers of smoldering
Sweethearts, flashes of citrus, stewed embers, hints of coastal seaweed, and a
apple, and growing smoke. The gentle, subtle sweetness. The finish is
glossy, mouth-coating finish generates exceptionally long, with a lingering
an urge to expertly puff smoke rings smokiness that warms the soul. At
out through pursed lips. (LIWC scores: 52.1%, it's a powerful testament to the
embodied cognition = 20; positive art of peated whisky, showcasing depth
emotion = 0) and character that only decades of
aging can achieve. A must-try for peat
enthusiasts. (LIWC scores: embodied
cognition = 13.04; positive emotion =
2.17)

Author Accepted Manuscript

Web Appendix C: Stimuli in Study 2 (Branded and Unbranded Content)

<p style="text-align: center;">Human copywriter (Branded content)</p>  <p style="text-align: center;">1,756 likes</p> <p>godiva Make your favorite homemade ice cream with this delectable topper. GODIVA buttery biscuits add the perfect crunch to your smooth and creamy ice cream.</p> <p><small>View all 16 comments</small></p> <p><small>6 hours ago</small></p>	<p style="text-align: center;">ChatGPT 4 (Branded content)</p>  <p style="text-align: center;">1,756 likes</p> <p>godiva Indulge in GODIVA's creamiest chocolate ice cream, sprinkled with decadent chunks. Every scoop is a piece of heaven. Truly a dream for chocolate lovers.</p> <p><small>View all 16 comments</small></p> <p><small>6 hours ago</small></p>
<p style="text-align: center;">Human influencer (Unbranded content)</p>  <p style="text-align: center;">5,038 likes</p> <p>brownedbutterblondie Not much can beat a Chocolate Cherry Cheesecake for one! You get a fudgy brownie base, silky smooth chocolate cherry cheesecake filling and rich chocolate ganache all to yourself with these mini treats! Topped with freshly whipped cream and a juicy cherry, these will put you in flavor heaven!</p> <p><small>View all 129 comment</small></p> <p><small>2 days ago</small></p>	<p style="text-align: center;">ChatGPT 4 (Unbranded content)</p>  <p style="text-align: center;">5,038 likes</p> <p>brownedbutterblondie Experience the delight of this exquisite chocolate cupcake. The moist cake base harmonizes perfectly with the luscious, creamy frosting. A drizzle of glossy chocolate sauce and a sprinkle of chocolate shavings add layers of richness. The crowning glory, a ripe red cherry, introduces a burst of freshness.</p> <p><small>View all 129 comment</small></p> <p><small>2 days ago</small></p>

Web Appendix D: Propositions for Further Research**Table W3.** Propositions for Further Research on Consumer Perception of AI-Generated Content

	Key Issues	Research Questions	Implications
1. The influence of product category on the characteristics of AI-generated narratives	<p>Can this research findings be extended to different product categories than the highly sensorial food and beverage category?</p> <p>Future studies should investigate a broader array of product categories along the utilitarian/hedonic continuum. They should examine how the narratives by generative AI perform against human-generated narrative in these different contexts.</p>	<p>How does the type of experiences and product category affect the differences between AI versus human generated narratives?</p> <p>What types of textual dimensions should be used to evaluate AI generated content for more utilitarian experiences?</p>	<p>Our findings should extend to other type of experiences that involve some hedonic dimensions such as restaurants, hotels, entertainment venues, cultural products, etc. The comparison between human and AI contents should reveal less differences for utilitarian products, such as home appliances or telecommunication services. These usage experiences involve more factual descriptions and evaluations (i.e., less affect), less bodily sensations (i.e. embodied cognition) and probably less variations from one consumer to the other (i.e., lower lexical diversity).</p>
2. Method to evaluate AI-generated marketing content	<p>In this research, two methods are used to compare AI versus human-generated contents: automated textual analysis and ratings by human judges. The automated textual analysis provides an objective and quantitative assessment of the texts. The results may sometimes diverge from the subjective perceptions by a human audience. The ratings of AI-generated content by</p>	<p>What is the appropriate method to measure AI-generated marketing content targeted to a consumer audience?</p>	<p>Future research in generative AI could develop refined text analysis tools that would replicate human evaluations. This might involve using generative AI to substitute for human participants in market research.</p>

Author Accepted Manuscript

	humans require large samples of participants, which is time-consuming and costly.		
3. The effect of prompts on the AI-generated marketing content	The prompts used to generate narratives in this research were generic and did not specify the specific dimensions on which the output would be evaluated. The prompts could be more explicit about the desired perception and effects on the reader to customize the LLM outputs.	How can AI-generated experiential narratives be improved by refining the prompts?	Future studies could instruct ChatGPT to produce texts with desired levels of embodied cognition, affect, and lexical diversity. A more fine-grained approach to prompts could also test the adaptability of LLMs to adjust the generated narratives to target audiences with specific sociodemographics, or level of expertise in the product category.
4. AI-generated audio and visual contents	This research focused on textual contents exclusively. Further investigation could probe the capacity of generative AI tools to create other types of contents (audio, visual) and how they compare to human generated similar content.	How do consumer evaluate audio and visual content generated by AI?	Future studies could investigate the implications of multi-modal AI-generated content applied to virtual influencers. Their “human-likeness” could be tested against human influencers. Deep fakes videos are another application context of non-textual AI-generated content.
5. The effect of AI authorship disclosure on consumer responses	Our research question examines how consumers perceive AI-generated experiential narratives without any hint about the AI authorship. Given the ethical issues of spreading potential misinformation, and the EU guidelines about transparency, further research could extend the investigation by manipulating the presence of an AI warning label.	How do consumers perceive and respond to the divulgation of AI authorship in marketing narratives?	Future studies could conduct experiments where the AI-generated content would include a disclosure label: the timing of the disclosure (before, during or after processing the narrative), and the content of the label (AI full or partial authorship, or the mention that there was no AI intervention) could be manipulated. The goal would be to test if the knowledge of the AI intervention will taint consumer evaluation of

Author Accepted Manuscript

			the generated narrative, and potential downstream consequences on attitude and engagement. This can provide guidelines for marketers to best implement public policy guidelines about transparency while still maximizing the intended effect of the marketing communication.
--	--	--	---

Peer Review Version