Journal of Advertising

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ujoa20

The MADE Framework: Best Practices for Creating Effective Experimental Stimuli Using Generative AI

Zeph M. C. van Berlo, Colin Campbell & Hilde A. M. Voorveld

JOURNAL OF ADVERTISING
2024, VOL. 53, NO. 5, 732–753
https://doi.org/10.1080/00913367.2024.2397777

The MADE Framework: Best Practices for Creating Effective Experimental Stimuli Using Generative AI

Zeph M. C. van Berlo[a], Colin Campbell[b], and Hilde A. M. Voorveld[a]

[a]University of Amsterdam, Amsterdam, The Netherlands; [b]University of San Diego, San Diego, California, USA

ABSTRACT
This paper introduces the MADE (Mapping, Assembling, Demonstrating, Executing) framework, a comprehensive set of best practices for the ethical and effective use of generative artificial intelligence (AI) in creating experimental stimuli for advertising research. The framework was developed through an extensive exploration of various emergent generative AI tools used in common experimental manipulations. We apply the MADE framework to demonstrate the creation of high-quality, realistic experimental ads using leading generative AI tools. Our empirical testing shows that AI-generated stimuli are valid, with consumers rating them equally high in quality, appropriateness, and realism compared with professionally created ads. This finding underscores the viability of AI-generated ads in advertising research. Additionally, we discuss the importance of adhering to ethical standards and ensuring transparency in AI use. By combining technological innovation with methodological rigor, this paper aims to guide researchers in leveraging the potential of generative AI while addressing its ethical implications, thereby enhancing the realism and validity of experimental advertising research.

The use of artificial intelligence (AI) in research is a popular topic, especially as related to the application of generative AI for content creation such as writing (e.g., Azrout et al. 2024; Bockting et al. 2023; European Commission 2024; Miao and Holmes 2023). These discussions often revolve around concerns over research quality, integrity, and copyright issues, leading editorial teams and associations like the American Psychological Association (2023) to develop guidelines for the appropriate use of generative AI in research. Despite the significance of these concerns and associated risks, an often-overlooked aspect of generative AI is its remarkable potential to aid researchers in developing experimental stimuli. This potential highlights an important area for exploration, suggesting that the conversation around generative AI in research should not only address ethical and procedural concerns, but also recognize its value in potentially enhancing research methodology.

Generative AI can automatically produce vast quantities of text, images, and videos from human-generated prompts (Peres et al. 2023). Public discussions often focus on its proficiency in crafting text nearly indistinguishable from human-written content. Currently, the most well-known platforms include the OpenAI ChatGPT, Google Gemini, and Microsoft Copilot platforms, along with rapidly evolving applications for generating images (e.g., DALL·E, Midjourney, Ideogram, and Artbreeder) and videos (e.g., Gen-2, Sora, Veo, RunwayML, and Synthesia). Additionally, specialized tools like Anyword and Copy.ai, designed to assist advertising professionals with tasks like copywriting, represent significant advancements in the ad industry. The academic community is also increasingly

CONTACT Zeph M. C. van Berlo Z.M.C.vanBerlo@uva.nl Amsterdam School of Communication Research (ASCoR), University of Amsterdam, P.O. Box 15791, 1001 NG Amsterdam, The Netherlands

Supplemental data for this article can be accessed online at https://doi.org/10.1080/00913367.2024.2397777.

Zeph M. C. van Berlo (PhD, University of Amsterdam) is an assistant professor of persuasive communication, Amsterdam School of Communication Research (ASCoR), University of Amsterdam.
Colin Campbell (PhD, Simon Fraser University) is an associate professor of marketing, Knauss School of Business, University of San Diego.
Hilde A. M. Voorveld (PhD, University of Amsterdam) is an associate professor of persuasive communication, Amsterdam School of Communication Research (ASCoR), University of Amsterdam.

examining the impact of generative AI on advertising and marketing sectors (e.g., Campbell et al. 2020, 2022b; Hayes et al. 2021; Rivas and Zhao 2023; Rodgers 2021; Thomas and Fowler 2021; Vakratsas and Wang 2021; Van Esch, Cui, and Jain 2021). Recently, Huh, Nelson, and Russell (2023) emphasized the need for further investigation into the application of AI within advertising research, highlighting the growing interest and potential implications of this new technology for the field.

Surprisingly, academic research is only starting to use generative AI for advertising research, particularly in experimental studies (e.g., Arango, Singaraju, and Niininen 2023; Sands et al. 2022). A fundamental challenge with experimental research is the creation of realistic stimulus materials, a task highlighted in Geuens and De Pelsmacker's influential article published in the *Journal of Advertising* 2017 thematic issue on advertising research methodology. They assert that experimental advertising stimuli must balance realism with experimental control, requiring that stimuli be professionally crafted, high quality, and convincingly realistic—essentially, indistinguishable from real advertisements. However, they also emphasize a recurring issue: The realism of ads produced by academics often falls short. This concern, reiterated by De Pelsmacker in 2021, is shared by other authors and editorial teams (e.g., Hamby and Russell 2023; Moorman et al. 2019; Morales, Amir, and Lee 2017; Van Heerde et al. 2021).

In the current paper, we demonstrate how generative AI can address the challenge of creating realistic and valid experimental stimuli for advertising research. Our principal contribution is the development of the MADE (Mapping, Assembling, Demonstrating, Executing) framework, which provides comprehensive guidelines for the responsible and effective utilization of generative AI. This initiative directly responds to the call by Huh, Nelson, and Russell (2023) for scholarly examination into responsible use of AI in advertising research. We articulate a set of best practices for developing stimuli using generative AI, ensuring that these recommendations are broad enough to cover various advertising contexts without being tied to specific AI technologies.

To provide insight into how generative AI can be leveraged by advertising researchers, this paper presents five steps. First, we overview what AI is and how it works, before turning to describe recent advances in generative AI. We do a similar review of the challenges posed in creating suitable stimuli for research and the various criteria used to assess stimuli. Second, we describe the process we used to develop our best practices framework to identify which generative tools exist, assess their effectiveness, and develop a deeper understanding of the process of using them to create stimuli. Third, we distill our insights into a comprehensive multi-stage best practices framework that we term the *MADE framework* and describe it in detail. Fourth, we apply the MADE framework to create stimuli for three types of manipulations common in advertising research. We then empirically assess the effectiveness of the generated stimuli in a repeated measures study. Finally, our conclusion offers reflections on the study, presenting a set of helpful practical recommendations for researchers starting out with incorporating generative AI into stimulus development, and we discuss the trajectory of generative AI and its broader repercussions for advertising research. Our comprehensive approach aims not only to guide current practices but also to inspire future innovations in the use of generative AI for advertising research.

Background

Generative AI Introduces New Opportunities in Advertising

Artificial intelligence refers to computer models that can perform tasks such as learning, reasoning, problem-solving, perception, prediction, and language understanding. Of advertisers worldwide, half currently use AI in some capacity (Stam 2023). *Generative artificial intelligence* refers to a subset of AI technologies that are designed to create new content in forms, such as text, images, videos, and more. Creation is through sophisticated algorithms and models, including *generative adversarial networks* (GANs) and transformer models like GPT (a generative pre-trained transformer). These models analyze vast datasets to learn how to produce content that is both like the existing examples it has learned on, but that is also distinct and original.

Generative AI works through a dual process in which an AI model (the *generator*) first generates content, and then a second model (the *discriminator*) evaluates it using GANs or GPT-based cycles of self-improvement that refine output using feedback loops (Campbell et al. 2022a). This iterative process is able to rapidly produce content that is often both high quality and relevant.

By providing new opportunities to engage audiences in a more personalized, efficient, and creative manner, generative AI has potentially major effects for advertising and marketing (Campbell et al. 2022a;

Chui et al. 2023). This perspective is echoed by advertising and marketing professionals themselves (Deveau, Griffin, and Reis 2023), who already employ a wide variety of tools aimed at different aspects of advertising. For instance, platforms like Jasper.ai and Writer.com use generative AI to produce creative and engaging text copy for ads. These tools leverage advanced natural language processing (NLP) techniques to understand context and produce copy relevant to target audiences, streamlining the content creation process.

Generative AI tools hold the promise of significantly reducing the time and cost associated with traditional content creation methods (Campbell et al. 2022a; Chui et al. 2023). These tools can also be leveraged for personalization and optimization, helping to potentially enhance campaign performance, drive higher engagement rates, and boost advertising return on investment (ROI). McKinsey (Chui et al. 2023) specifically highlights advertising and marketing as areas with considerable potential to derive value from advancements in AI technology. In addition to the new opportunities for generative AI in advertising, we believe generative AI might also offer opportunities for advertising research.

Navigating the Realism–Control Paradox in Experimental Stimulus Design

In recent years, numerous authors and editorial teams have expressed concerns over the realism of advertising and marketing stimuli, arguing it limits the practical relevance of advertising research to professionals in the field (De Pelsmacker 2021; Hamby and Russell 2023; Moorman et al. 2019; Morales, Amir, and Lee 2017; Van Heerde et al. 2021). In their seminal paper, Geuens and De Pelsmacker (2017) provide comprehensive guidelines for the planning and execution of experimental advertising research, covering aspects from sampling to questionnaire design. Importantly, they emphasize methodological best practices for stimulus development, advocating for a delicate balance between realism and control. This balance, they argue, is crucial in developing experimental advertising stimuli that are both effective and scientifically rigorous.

As a starting point, advertising stimuli should be professionally crafted, high quality, and realistic, essentially mirroring genuine advertisements (Geuens and De Pelsmacker 2017; De Pelsmacker 2021). This requirement is important because the advertising landscape and its contextual backdrop are continually evolving. However, realism is not the sole consideration. Experimental designs often involve the manipulation of a specific factor, and it is imperative that this manipulation is executed validly and with precision.

A common guideline is to alter only the necessary elements for the manipulation while maintaining uniformity across other aspects of the advertisement (Geuens and De Pelsmacker 2017). Optimal manipulation typically involves simplifying or minimizing complexity in the rest of the material, to avoid introducing any potential confusion. This necessity to balance professional realism with clear and concise manipulation introduces what is termed a *paradox* between control and realism. It highlights the challenge in creating stimuli that are both authentic and valid.

Generative AI: A Catalyst for Improving Validity Measures in Advertising Research?

In evaluating the quality of research, essential criteria include *construct validity*, *statistical conclusion validity*, *internal validity*, and *external validity*, and its subset, *ecological validity*. These elements are often considered together, because potential tradeoffs among them are frequently present (Hamby and Russell 2023). This paper argues that generative AI holds significant promise for enhancing the construct validity, internal validity, and external validity as well as the ecological validity of experimental advertising research. By leveraging generative AI, it is possible for researchers to achieve more nuanced, realistic, and controlled experimental designs, likely improving the overall robustness and value of their findings.

Construct validity is defined as the degree to which research operations (e.g., manipulations) can be generalized to higher-order constructs, essentially questioning whether the researchers manipulated the construct as intended (Hamby and Russell (2023) based on Banaji and Crowder (1989) and Shadish, Cook, and Campbell (2002)). To assess construct validity, researchers typically conduct manipulation checks. Generative AI offers a novel method for enhancing construct validity by enabling the manipulation of ad elements that are traditionally challenging to modify. For instance, adjusting the age of a model depicted in an advertisement—a task that is usually cumbersome for researchers to achieve manually—becomes feasible with generative AI.

Internal validity measures the confidence that a cause-and-effect relationship observed in a study is not influenced by external, confounding factors

(Perdue and Summers 1986; Hamby and Russell 2023). Generative AI has the potential to contribute by enabling precise manipulation of a specific element within an advertisement while ensuring all other elements remain unchanged. For example, it allows for the modification of product type or the facial expression displayed by a model in an advertisement without altering any other aspect of the ad. Doing so maintains the integrity of the experimental design and bolsters internal validity.

External validity is defined as the "validity of conclusions about the generalizability of a causal relationship to and across populations, of persons, settings, and times" (Hamby and Russell 2023, 2). It can, for example, be tested by replicating the effects of one experiment with another ad, brand, or population. Generative AI can facilitate this replication because it makes it easier to design multiple versions of a stimulus, such as embedding the manipulation in different specific designs or layouts. This approach may increase the robustness of findings and theories.

Ecological validity evaluates how well the effects observed in a laboratory setting translate to real-world scenarios (Hamby and Russell 2023). This evaluation involves determining whether the experimental manipulations are perceived as realistic and representative of actual advertisements encountered in everyday life. Generative AI holds considerable promise in this area, offering the ability to produce advertisements that closely mimic those seen in professional settings. This capability potentially addresses the common challenge of experimental advertising stimuli appearing unrealistic compared with professionally developed ads (De Pelsmacker 2021).

Finally, *statistical conclusion validity* refers to the accuracy of conclusions drawn about relationships among variables based on the data (Hamby and Russell 2023). This measure examines whether manipulations are potent enough to elicit changes in dependent variables. Traditionally, researchers were often limited to using preexisting advertisements, constraining their ability to design effective manipulations. However, generative AI enables the creation of distinctly different versions of an advertisement, specifically tailored to the desired manipulation. Concretely, it allows researchers to easily generate multiple versions of an ad that differs only in the independent variable. Ultimately, this approach offers greater control over the manipulation's effect size, by allowing the selection of AI-generated ads with either maximal or minimal variation in the independent variable—while all other elements of an ad remain constant.

Developing Guidelines for Advertising Stimuli Creation Using Generative AI Tools

To develop the guidelines for advertising stimuli creation described in following text, we carefully explored key aspects of generative AI use. First, we sought to map the set of image-capable generative AI tools currently available to researchers. All three authors and two research assistants undertook independent searches for generative AI tools by searching the internet, research published in advertising journals, and articles in the popular or professional press. Results of individual searches were then combined, resulting in a list of 16 tools. This list shows that a distinction can be made between generative AI tools capable of either creating or editing.

After developing the list of tools, we assessed their performance by immersing ourselves in the process of using generative AI through first-hand experience and observation (Creswell and Poth 2016). During this process, we decided to focus on visual experimental stimuli rather than on textual, audio, or audiovisual stimuli for several reasons. First, visual stimuli are used most in recent publications in advertising journals (i.e., Hamby and Russell 2023 note that only 19 studies published in the *Journal of Advertising* between October 2019 and January/February 2021 used an audio or audiovisual format). Second, we quickly concluded that generative AI tools for video formats cannot deliver the quality needed when this paper was written (April 2024). Thus, the guidelines and best practices discussed in this paper can also be applied to stimuli with other modalities.

Two of the authors independently tried each of the tools that were identified. This approach involved, as was possible with each, testing out how each tool performed editing and creating images, and how the tools could be used to develop stimuli pairs by either editing an image to incorporate a specific difference, or by generating two new images that varied by the specific difference. Articles, video, and help resources were consulted to ensure each generative AI tool was being used to its full potential. Another key aspect in the process was testing multiple ways of specifying a prompt.

Following our investigations, six tools—three for editing (Adobe Photoshop Firefly, DreamStudio.ai, and Clickdrop) and three for creation (OpenAI DALL·E, Bing Image Creator, and Midjourney)—were selected as current industry leaders and marked for further investigation in our later empirical investigation. This selection was based on both each tool's capability to either edit or generate changes relevant to stimuli creation, as well as the resulting quality of the tool's output. In sum, following the pattern-matching technique outlined by Yin (1994), we first individually characterized our respective processes before sharing them to identify overarching themes and develop a comprehensive framework for use of generative AI to develop stimuli. We describe this model in the next section.

The MADE Framework for Effective Generative AI-Based Stimulus Development

In this section, we propose our MADE framework (see Figure 1) for stimulus development with generative AI. This framework consists of eight fundamental steps distributed across four phases—*Mapping*, *Assembling*, *Demonstrating*, and *Executing*—that enhance the prospects for effective application of generative AI in stimulus development. It is crucial to understand that although these steps are outlined sequentially for clarity, the process in practice frequently involves recursion and iteration, or using the words of McGrath "the process really should be regarded as a series of spirals" (McGrath 1981, 181). We now provide a more detailed description of each step below.

Mapping

Determine Manipulation

Variable manipulation is key in experimental design and essential for establishing causal links between variables (Shadish, Cook, and Campbell 2002). To determine cause and effect, researchers manipulate a presumed cause and then measure a resulting effect. Recent publications in the *Journal of Advertising* offer examples of such manipulation, including the portrayal of movement in advertisements (Grigsby, Jewell, and Zamudio 2023), the comparison between virtual influencers and human endorsers (Franke, Groeppel-Klein, and Müller 2023), and the exploration of ad creativity alongside interference (Lehnert et al. 2023).

Determine Context

Another key element of experimentation is context, which involves understanding the target audience, the advertising platform (e.g., social media, magazines, news websites), and the cultural or social nuances affecting ad reception. Context greatly influences generalizability to real-world settings (external validity). If the context is too artificial, findings may not apply outside the laboratory. Media contexts like social media feeds, magazines, or broadcast programs are particularly important. It is crucial to determine the context for embedding AI-generated ads or experimental stimuli in advance because this context can impact the selection of AI tools.

Choose Existing Ad (Optional)

Using AI tools to generate preliminary prompts is often more efficient and yields higher quality results than creating prompts from scratch. A practical method is to use an existing ad as a reference. For instance, Midjourney's "describe/" function can prompt AI to describe an uploaded image, which can then be used to create a new image based on its characteristics. Our experience is that this method tends to lead to the highest quality results.

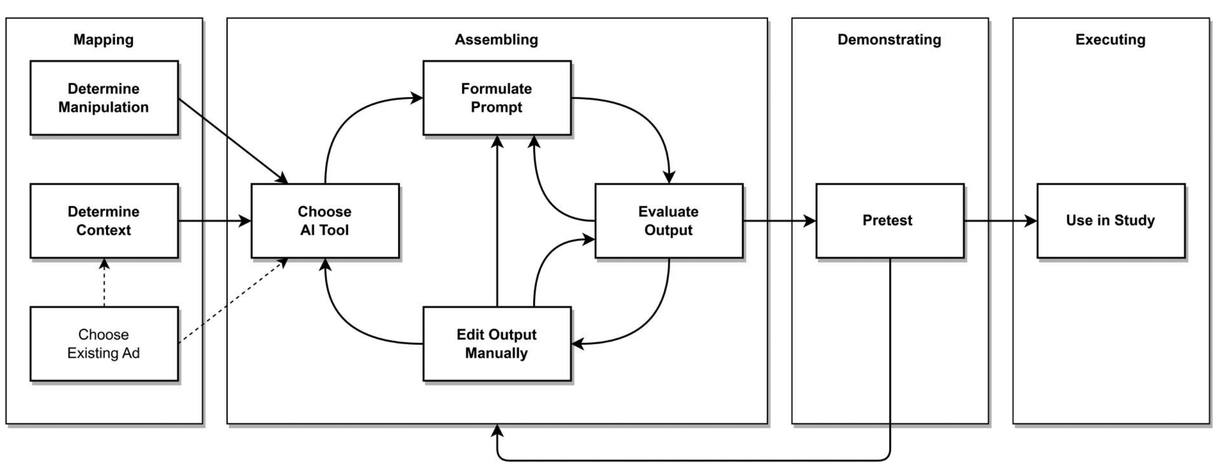


Figure 1. MADE (Mapping, Assembling, Demonstrating, Executing) framework for stimulus development with generative AI.

Assembling

Choose AI Tool

Selecting the appropriate tools to begin developing AI-generated content is critical. The decision should be based on the tool's capability to create a particular type of output. Currently (May 2024), different AI tools offer different types of output. In the advertising sector, AI tools are being utilized for a range of functions, including the following (with examples in parentheses):

- Prospecting (6sense.com)
- Data management and customer engagement (segment.com)
- Copywriting (Jasper.ai and Writer.com)
- Search engine optimization (SEO) (SurferSEO.com, Headlime.com, Semrush.com)
- Chatbots (Chatfuel.com, Drift.com)
- Digital advertising personalization and optimization (albert.ai),
- Media monitoring and competitive intelligence (Brand24.com, Crayon.co)
- Influencer selection (influencity.com)
- Social media engagement (Manychat.com)
- Public relations campaigns (Howler.media)
- Social media content creation (predis.ai, Flick.social)
- Visual content creation (DALL·E, DreamStudio.ai, midjourney.com, RunwayML, Synthesia, Artbreeder)
- Writing advertising texts (ChatGPT 4o)

Many of these generative AI tools are standalone applications, focusing on a particular type of output. For example, the original *ChatGPT 3.5* focused on generative text content, whereas *Midjourney* focuses on generating images. It is conceivable that AI tools will evolve into "one-stop-shops" for all types of AI-generated content. This trend is already evident in the OpenAI ChatGPT platform, which began as a text-focused AI tool but integrated the generative image AI, DALL·E, in 2023. Now, this platform can generate both textual and visual output.

As new tools emerge and existing ones are phased out, selecting the right AI tool requires a deliberate approach. Researchers should first identify the tool type based on its intended purpose (e.g., copywriting, visual content development) and then carefully evaluate the pros and cons of each tool within that category. Table 1 summarizes the leading tools currently available, although this landscape will continue to evolve. In some cases, using a combination of tools may be necessary to achieve the desired outcomes.

Formulate Prompt

Prompt engineering, or crafting inputs to guide generative AI models, combines creativity with technical knowledge to produce desired outputs. It is a precise and strategic approach requiring researchers to anticipate AI responses and adjust prompts for optimal results. Carefully crafted prompts enhance the relevance and accuracy of AI-generated content. The process of prompt engineering is iterative. It starts with creating a prompt, generating output, and refining the prompt to better align the generated content with the intended outcome. Starting with a broad prompt and progressively narrowing its focus helps direct the AI more accurately (Pot 2023). General guidelines for prompt formulation suggest using active language and concrete nouns to create dynamic and precise inputs (Pot 2023).

Including detailed context information is vital for generating high-quality AI content. The more context provided, the more relevant the output. Without sufficient context, AI outputs may be too generic or miss important nuances intended by the researcher. Contextual details can include advertising goals, brand identity, target audience specifics, cultural context, creative direction, and media format. These details directly impact the relevance and utility of the AI-generated output, making them essential for achieving desired outcomes (Pot 2023).

Additionally, clearly describing the preferred execution of the material is beneficial. This approach includes outlining limiting factors such as specific colors, objects, writing style, perspective, and tense. Defining the focal point and desired composition—whether a close-up, wide shot, or bird's-eye view—is also crucial. Specifying the ambiance or emotion intended for the content ensures that AI-generated output aligns with the

Table 1. Overview of assessed AI tools.

Primary Use	Name	Image Quality	Speed	Ease of Use	Price
Image creation	OpenAI's DALL·E	High	Moderate	Moderate	$20 US monthly
Image creation	Bing Image Creator	Moderate	High	High	Free
Image creation	Midjourney	High	Moderate	Moderate	US$10 monthly (basic plan)
Image editing	Adobe Photoshop Firefly	High	Moderate	Low	US$34.49 monthly
Image editing	DreamStudio.ai	Moderate	High	High	Approx. US$10 per 100 images (cost depends on image size)
Image editing	Clickdrop	Moderate	High	High	US$15 monthly

Note. The prices reflect those in August 2024.

campaign's emotional tone and atmosphere (Pot 2023). Given that outputs can vary significantly between AI tools (an example is included in Figure 2), trying out prompts across several tools is recommended. This comprehensive approach to specifying preferred execution increases the likelihood of meeting objectives and effectively resonating with the target audience.

Evaluate Output

After the AI tool has been used to manipulate the advertisement, the output is evaluated to ensure it meets the desired criteria. This evaluation can for example be based on esthetic quality, realism of the elements, alignment with marketing goals, or potential effectiveness. Some examples of common generation errors (e.g., uncanny hands/faces, disproportionate elements) to look out for are shown in Figure 3. When satisfied with the output, researchers are recommended to conduct a pretest before using the materials as stimulus materials (see Pretest step). However, often, AI-generated images might need some minor manual editing before they are ready to be used in a study.

Edit Output Manually

Many generative AI tools allow users to interact with and modify the content after it is created, enabling an iterative creation process. However, manual editing of the generated materials using software like Photoshop can sometimes be more efficient than continuously altering the prompt. For example, it is often necessary to manually add a brand name and slogan to an AI-generated ad because many AI tools avoid generating content that includes copyrighted materials such as logos and products. Manually adding or changing text, such as a slogan, is often more efficient with image editing software than by adjusting the prompt.

It is also important to note that text content, such as slogans, can be created using generative AI tools like ChatGPT. To increase the ecological validity of stimulus ads, following brand guidelines for logo and slogan placement is best practice. These guidelines, typically available on a company's website, provide relevant information on typography, colors, layout, suitable imagery, and tone of voice. Adhering to these guidelines ensures that experimental stimuli closely mimic real-world advertisements, thereby improving their ecological validity.

Demonstrating

Pretest

First, conduct a pretest before launching the main study to evaluate the suitability of experimental procedures and instruments (Perdue and Summers 1986). Pretesting AI-generated materials involves examining manipulation clarity, identifying and testing confounding variables, assessing ecological validity, and considering ethical implications.

It is critical to verify the manipulation with detailed assessments and quantitative checks (Perdue and Summers 1986; Hamby and Russell 2023). Confound checks ensure manipulations do not affect unintended variables (Cook and Campbell 1979; Perdue and Summers 1986). For instance, when using AI to change product type in ads, maintaining consistent ad quality is crucial. The next step is to assess ecological validity with qualitative and quantitative measures to ensure stimuli are high-quality, appropriate, and realistic.

Pretesting allows for necessary adjustments, ensuring effective and accurate stimuli for the main study (Hofer 2018; Perdue and Summers 1986). If significant confounds are found or stimuli are rated as unrealistic, return to the assembly phase for refinement. This feedback loop ensures the manipulation is

Figure 2. Output with same prompt across three AI tools. These images were generated/edited with artificial intelligence. From left to right: DALL·E 3, DreamStudio, Midjourney. The prompt that was used was: "a woman with a joyful expression, styled for a perfume advertisement on a beach, set against a backdrop of light orange and light magenta hues."

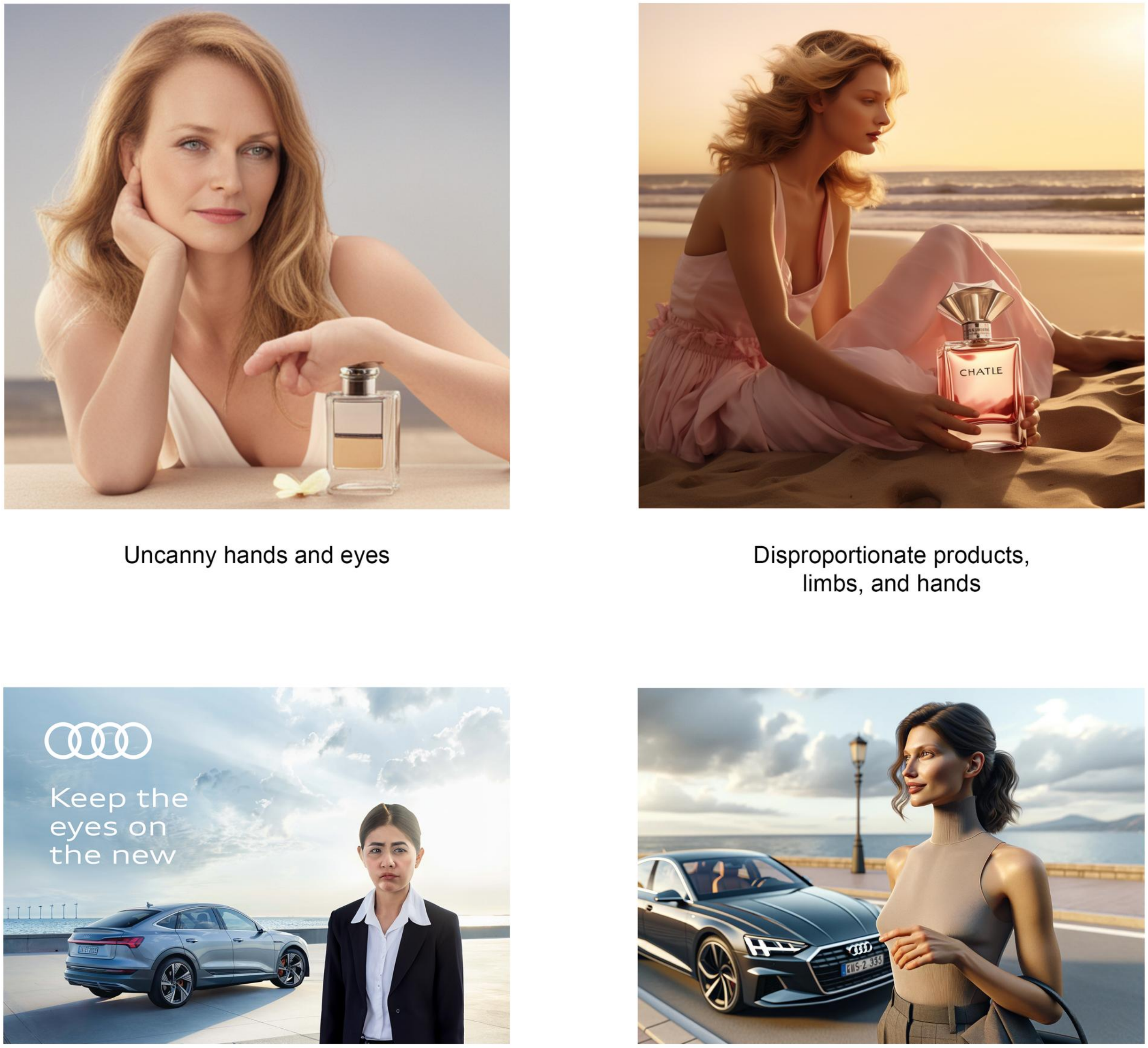


Figure 3. Several common AI-generation errors. These images were generated/edited with artificial intelligence: DreamStudio (top left), Midjourney (top right), Adobe Firefly (bottom left), and DALL·E 3 (bottom right).

perceived as intended and the final experimental materials are of high quality.

Executing

Use in Study

If the pretest results are positive, use the refined stimuli in the primary study. Readminister the main measures and checks (manipulation, confound, and ecological validity) as in the pretest phase.

Ethical Guidelines

In employing generative AI as a research tool, researchers retain ultimate responsibility for its application within the scientific process. This responsibility includes thoroughly assessing ethical considerations, including addressing potential AI biases, checking generated material for diversity and inclusiveness, navigating copyright concerns, understanding the necessity for transparent disclosure, and adhering to country-specific legal frameworks related to AI usage. Given the large variability in institutional and national guidelines, we recommend that researchers consult with their local institutional review boards, data stewards, privacy officers, and potentially even legal advisors early in the research planning process because having many actors involved can slow down the research process considerably. Additionally, researchers should stay informed about ongoing developments in responsible AI use guidelines (e.g., Bockting et al. 2023; European Commission 2024; Rivas and Zhao 2023). Such a comprehensive approach ensures that the study not only respects ethical standards but also reduces legal and societal risks associated with AI-generated content.

Disclosing the Use of Generative AI

When ethically using generative AI, it is essential to recognize that employing AI-generated images to portray an unreal scenario may be perceived as deceptive and currently strong societal concerns exists about how generative AI can function as "the ultimate disinformation amplifier" (e.g., Endert 2024). Now regulations are being developed that prescribe that content generated by AI should be labeled as such. For example, the AI Disclosure Act (U.S. Congress. House 2023) in the United States requires that a disclaimer stating, "This output has been generated by artificial intelligence," be added to content generated using AI. Also in relation to discussion on digital literacy, it is believed that consumers, and vulnerable groups in particular, have a right to know about the technology behind the ads they encounter (Voorveld, Meppelink, and Boerman 2023; Wang et al. 2024; Wu, Dodoo, and Wen 2024). Therefore, a debriefing session at the study's conclusion to maintain ethical standards is needed.

The importance of disclosing the use of generative AI tools in research underscores a commitment to ethical research practices. This approach aligns with the recommendations set forth in the *living guidelines for generative AI* by Bockting et al. (2023), which advocate for full transparency regarding the use of generative AI in research endeavors. Adhering to these guidelines not only fosters trust but also aligns with the *Journal of Advertising* principles for research quality and ethics. As outlined by Huh, Nelson, and Russell (2023), transparency and responsibility are cornerstone principles for the ethical use of generative AI in research.

Copyright

Whether the use of generative AI in general infringes on the copyright of those whose content is used for training is an ongoing debate (European IP Helpdesk 2023; Yoon and Borges 2024). However, the use of copyright protected materials for research is allowed in many parts of the world under certain provisions. For example, in the United States, the *fair use* legal doctrine allows limited use of copyrighted material without requiring permission from the rights holders for purposes such as commentary, education, and research. When using copyrighted material for research purposes, however, it is considered best practice to include a reference to the original source. Also, to avoid any potential copyright-related issues, it is recommended to use public domain materials if these are available.

Privacy

Besides transparency and copyright, also privacy protection is important to consider when using generative AI for research. In this context, it is critical to remember that information shared with AI tools is typically stored by the service provider and could be used for other purposes, such as training AI models. We therefore caution against uploading sensitive, confidential, or personally identifiable information without taking any mitigation measures, to avoid it being processed by and integrated into the AI model. Mitigation measures could include anonymizing data before using it, or, in case these data are available, making sure the settings of the AI tool restrict the training based on one's input. Notably, once data are used to train an AI model, it is currently impossible to remove the data from the model (Miao and Holmes 2023)—and, in extreme cases, it is even believed other users can retrieve it (Cohen, Evgeniou, and Husovec 2023). Therefore, before using any AI tool, it is essential to review its terms of service and understand the implications for any uploaded information.

An Empirical Illustration and Test of the MADE Framework

After developing and presenting the MADE best practices framework, we now turn to showing how the MADE framework can be applied. We do so by first using the MADE framework to develop stimuli pairs using generative AI. We then empirically test consumer responses toward the generated ads to quantitatively assess their performance in terms of the key forms of validity identified earlier.

Mapping

Determine Manipulation

A review of experiments published in four recent issues of *Journal of Advertising*—along with our understanding of the present capabilities of the generative AI tools—led us to explore three potential interesting manipulations: A change in age (younger vs. older), a change in the facial expression of the model featured in the ads (smile vs. neutral), and a change in product (either a change in product type or change in the color of the featured product).

Determine Context

In terms of context, we decided to investigate both more traditional-looking visual advertisements that could appear in an (online) magazine, newspaper, or on social media, as well as a social media influencer post. To at least somewhat improve the external validity of our findings, we decided to implement all three manipulations into different types of advertisements.

Choose Existing Ads (Optional)

Three existing images were then selected for testing: a car ad (Audi 2020), a perfume ad (Louis Vuitton 2018), and a social media influencer's sponsored post for cereal (brihallofficial 2018). We decided to use nonpublic domain images here because the aim of this research was to show how generative AI advertisements compare with existing advertisements. Then, we decided to implement our manipulations according to the following plan. In the case of the car ad, this manipulation was changing the model's expression and age and changing the color of the car. For the fragrance ad, we again changed the model's expression and age, as well as the product from a fragrance bottle to an apple. For the social media influencer post, we changed model age and expression, as well as the product from a cereal box to a fragrance bottle.

Assembling

Choose AI Tool

As described earlier, our exploration of generative AI tools led to six tools-three for editing and three for creation-being selected as industry leaders based on their ability to either edit or generate changes relevant to stimuli creation, as well as the resulting quality of its output. To be comprehensive, we decided to attempt creation using of the nine pairs of stimuli using all six generative AI tools. This approach necessitated that 56 pairs of images (3 changes for each of 3 images, created using each of the six tools) were independently attempted by two of the authors.

Formulate Prompt

Two authors independently created prompts for use with each generative AI tool. In most cases, this prompting was a recursive process that involved revisions to try and overcome limitations or problems with a particular tool's output.

Evaluate Output

Following creation of the stimuli pairs, we again met to compare experiences and discuss our findings with the third author who was not involved in creating the stimuli. For each specific manipulation of a particular image, we examined all of the image pairs we created. Examples of output that did not meet our final selection can be found in the Supplemental Online Appendix (Figures A1–A3). Interestingly, the strongest stimulus pairs were all created by a single tool: Midjourney.

Edit Output Manually

After making the final selection of our 18 generative AI-created stimuli pairs (calculated as 3 manipulations, all with 2 versions (= 6) × 3 different contexts), images were manually edited to include brand names, products, and slogans. In Table 2, we included a step-by-step description of the steps in the Mapping and Assembling stages for one pair of ads as an example.

Demonstrating

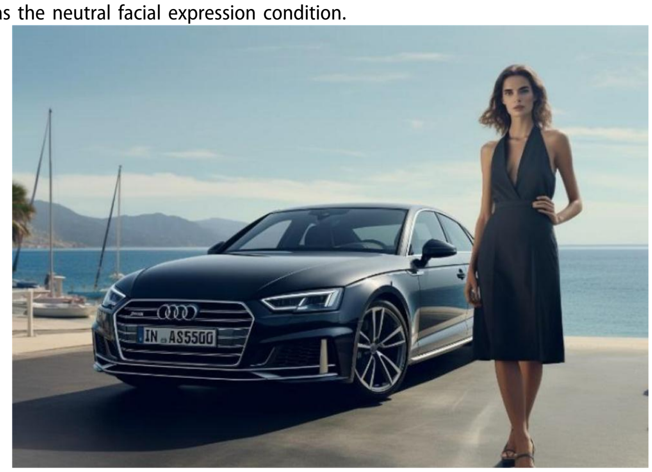
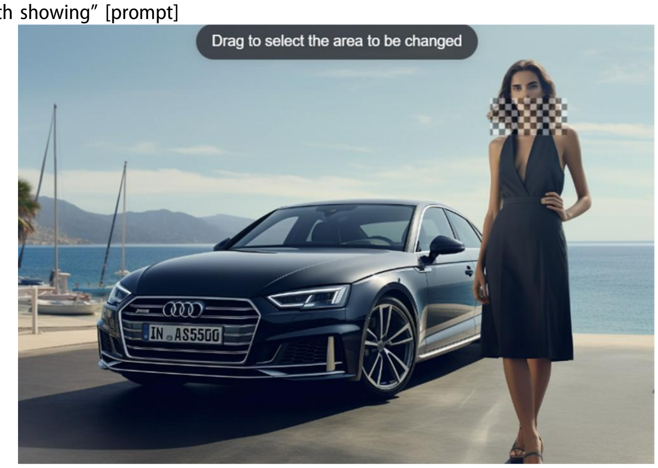
Pretest

Because the aim of our empirical illustration was to show how the MADE framework can be implemented and to show that it helps to develop valid advertising stimuli, we performed a quantitative pretest to assess the relevant forms of validity identified earlier in the paper. This study received ethical approval from the ethical review board of the Department of Communication Science, University of Amsterdam (reference number: FMG-5405). To aid in our comparison, we chose to also include for testing the three original, professionally created images (i.e., Audi 2020; brihallofficial 2018; Louis Vuitton 2018) that the AI-generated stimuli were based on. This approach brought the total number of images being tested to 21 (9 pairs [= 18], plus the 3 original, professional images). The ads used are in Tables 3–5.

Sample and Design

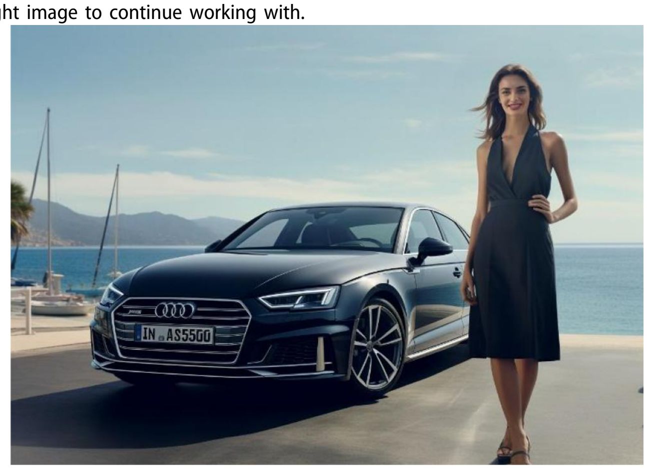
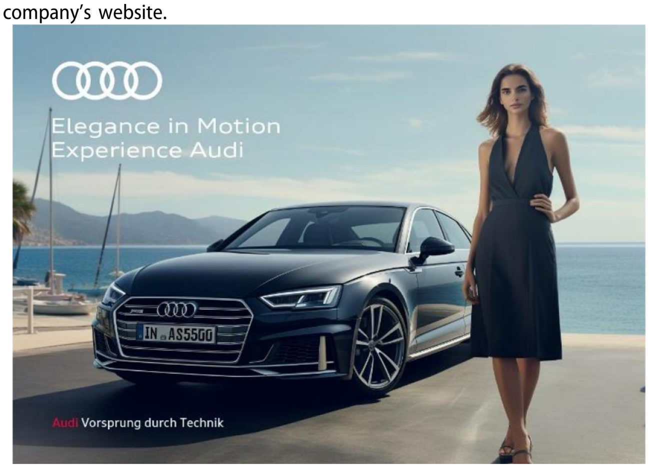
This study used Prolific Academic to recruited 402 participants from the United States. The average age was 40.7 years ($SD = 1.8$), 50.2% were female, and the highest levels of education acquired were high school (29.0%), bachelor's degree (44.3%), and master's degree (17.3%). The study took an average of 18 minutes to complete, and participants were paid

Table 2. Example case mapping and assembling phases of the MADE framework.

Step	Description
Determine manipulation	Expression (smiling vs. neutral)
Determine context	The context was a car advertisement. We used an existing ad as a starting point by using Midjourney's "/describe" function to generate a context prompt. We received four potential prompts to use as context, of which we chose: "A woman standing in front of a car with the words keep the eyes on the new Audi, in the style of seaside vistas, hazy, the new fauves, energy-charged, polished craftsmanship, dom qwek,[a] 8k resolution"
Choose AI tool	Midjourney
Formulate prompt	We started generating by using the context as a prompt.
Input	"A woman standing in front of a car with the words keep the eyes on the new Audi, in the style of seaside vistas, hazy, the new fauves, energy-charged, polished craftsmanship, dom qwek, 8k resolution" [prompt]
Output	
Evaluate output	We chose the bottom right image to continue working with. Because the model has a neutral facial expression, this image will serve as the neutral facial expression condition.
Output	
Formulate prompt	To create the stimulus ad for the smiling expression condition, we used the output for the neutral condition as a starting point. We selected the bottom part of the model's face to adjust, and used the following prompt (input).
Input	"Smiling mouth teeth showing" [prompt] 

(*continued*)

Table 2. Continued.

Step	Description
Output	
Evaluate output	We chose the top right image to continue working with.
Output	
Edit output manually	Once we had generated the images for both the stimulus ads, we used image editing software (i.e., Photoshop) to add the logo and slogan of the target brand to both stimulus ads. The slogan "Elegance in motion experience Audi" was created using the generative AI tool ChatGPT 4.0. To increase ecological validity, we followed the brand guidelines on the company's website.
Output	

(*continued*)

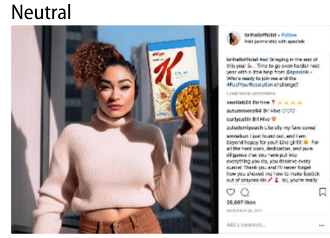
Table 2. Continued.

Step	Description
	
Evaluate output	The output was considered ready to pretest as stimulus ads. This readiness meant we moved on to the Demonstrating phase of the MADE framework.

Note. The process presented here is (for clarity) presented, for the most part, in a linear way. The process was actually more iterative, meaning that some steps were repeated multiple times to achieve the best results. The images included in this table were generated with artificial intelligence (Midjourney) and draw from the art style of Dominic Qwek (Dom Qwek n.d.).

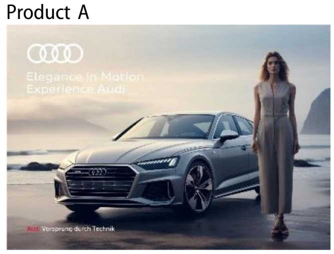
[a]The term "dom qwek" refers to a creature and character artist, Dominic Qwek, who has contributed to games such as StarCraft II, Diablo III, Overwatch and Killzone 2 (Dom Qwek n.d.). This term, when included in a prompt, signals that the generated image should draw from the art style of this artist. It is essential to acknowledge this approach when drawing from a specific artist's style of art.

Table 3. Overview influencer ads.

Type of Ad	Manipulation	Version A	Version B
AI generated	Age	Younger 	Older 
AI generated	Expression	Smiling 	Neutral 
AI generated	Product	Product A 	Product B 
Original ad	brihallofficial (2018)		

Note. The images included in this table were generated with artificial intelligence (Midjourney).

Table 4. Overview car ads.

Type of Ad	Manipulation	Version A	Version B
AI generated	Age	Younger 	Older 
AI generated	Expression	Smiling 	Neutral 
AI generated	Product	Product A 	Product B 
Original ad	Audi (2020)		

Note. The images included in this table were generated with artificial intelligence (Midjourney) and draw from the art style of Dominic Qwek (Dom Qwek n.d.).

the equivalent of US$5.75. One participant was excluded for taking an abnormally long time to complete the study (i.e., more than 50 times longer than the average time spent) . A CAPTCHA question was also used to block non-human participants (i.e., robots) from completing the study, which excluded one participant. Our final sample was 400 participants.

For this study, we adopted a within-subject experimental design, with each participant rating all of the tested advertisements. Although within-subject designs are often noted to have potentially lower external validity compared with between-subjects designs, they provide higher power and sensitivity in detecting differences between conditions (Schmidt and Hunter 2015). This feature is crucial for our study, given the subtle nature of the manipulations across 21 images, for which detecting nuanced differences is essential. Participants were told "the goal of the study is to form a clearer idea of how people respond to different images." To avoid bias, the word "images" was used throughout the study to refer to the advertisements being tested. However, some scale items did ask how the images compared with advertisements.

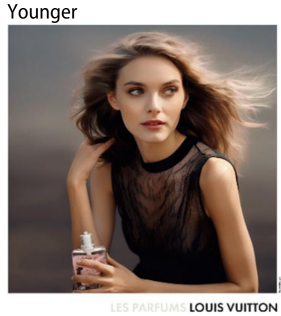
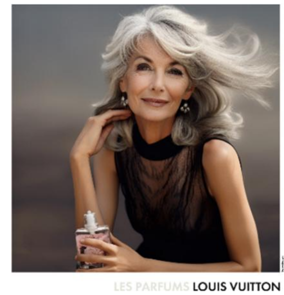
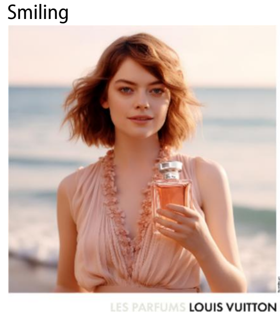
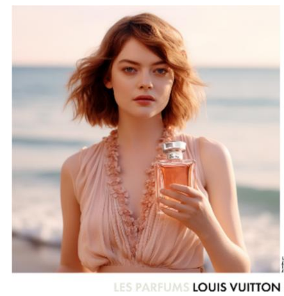
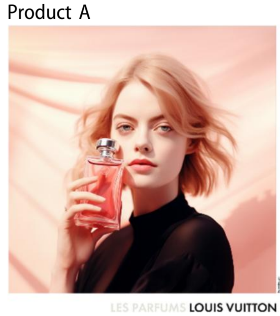
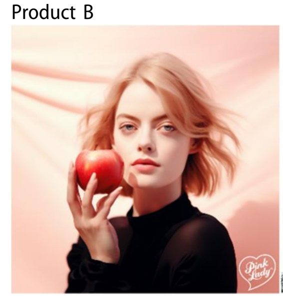
After completing informed consent, participants were then told, "In this study you will be shown a set of 21 images. For each image, you will be asked to give your first impression. Please note that there are no right or wrong answers, we are just interested in your honest first impressions." Participants then proceeded to have a brief look and report their first impressions of each of the 21 advertisements in a random order. The advertisements were shown in a random order to avoid order-effects. The participants' time spent for this step was recorded for each of the 21 advertisements. After completing the questionnaire, the participants were debriefed and told that 18 of 21 of the images they had evaluated were created using generative AI.

Measures

After viewing each advertisement, participants responded to eight items. These items assessed constructs representing the key forms of validity discussed earlier for which we believe generative AI could bolster when developing stimuli.

Construct validity was assessed by a one-item manipulation check for every manipulation we

Table 5. Overview fragrance ads.

Type of Ad	Manipulation	Version A	Version B
AI generated	Age	Younger 	Older 
AI generated	Expression	Smiling 	Neutral 
AI generated	Product	Product A 	Product B 
Original ad	Louis Vuitton (2018)		

Note. The images included in this table were generated with artificial intelligence (Midjourney).

included. With regards to the age manipulation (younger vs. older model), we asked: "Approximately, how old (in years) do you think the person in the image is?" For the facial expression manipulation (smile vs. neutral), we asked: "To what extent does the person in the image display happiness?" The answer option range was from 1 *not at all* to 7 *heavily present*. To assess manipulation of product type, we asked: "What product is the person holding in the image?" Answer options included *an apple, perfume, chips, cereal*, or *a smart speaker*. Finally, to assess the color manipulation, we asked: "What is the color of the car in the image?" with answer options *red, blue, silver, black*, and *gold*.

Internal validity was assessed using a one-item confound check and a time-spent measure. A possible confounding variable for all our manipulations was whether the attractiveness of the model or product changed when manipulating age, facial expression, product type, or color. To assess this variable, we asked participants to respond to the statement: "The image… shows an attractive looking product/shows an attractive model/shows a model that is attractive for her age." The answer options ranged from 1 *strongly disagree* to 7 *strongly agree*. We also assessed time spent on each image's questionnaire page viewing and responding to questions about the image. Our logic is that images attracting extra time are an indication of an image being abnormal or special in some way.

Ecological validity was assessed using six items focusing on three subdimensions of ecological validity: *quality*, *appropriateness*, and *realism*. An overview of the exact items can be found in Table 6. Statistical conclusion validity was not explicitly assessed because we did not compare the effects of manipulations using generative AI tools with traditional manipulations.

Table 6. Measurement items ecological validity scales.

Construct	Item	Reliability[a]
Quality	"The image is high quality."	.90
	"The image looks professionally created."	
Appropriateness	"The image looks like a [type of advertising]."	.61
	"The image is suitable for the product being marketed."	
Realism	"The image resembles a professionally designed advertisement."	.81
	"The image mimics the appearance of a real-world advertisement."	

Note. Spearman-Brown split-half reliability coefficients were estimated, rather than Cronbach's alpha, as recommended by Eisinga, Grotenhuis, and Pelzer (2013) for constructing two-item scales.
[a]Reliability estimates are the average reliability across 21 measurements.

Results

To evaluate the construct validity of our AI-generated ad pairs, we performed manipulation checks for all nine advertisement pairs. To test the age and expression manipulation, paired-samples t-tests were estimated. As shown in Table 7, we found significant differences between the first six pairs. All effect sizes were considered large (> 0.80). The models in the three younger ad groups were believed to be younger than the models in the three older ad groups. For the expression manipulation we found that the three smiling models were evaluated as being happier than the neutral-looking models. For the product manipulation, we asked participants to indicate which product the model was holding (or, in the case of the car, what color was the car). We found a near-perfect accuracy ($> 99\%$) for participants in identifying the products in the ads, apart from the color of the car manipulation. The correct color of the car was indicated in 92% of the cases. Overall, these results indicate that we successfully manipulated the age and facial expression of the model in the ad and the product type and support the construct validity of the AI-generated ad pairs.

To determine the internal validity of our AI-generated ad pairs, we compared the relative attractiveness of the models (for the age and facial expression manipulation) and products (for the product manipulation) for each of the nine pairs. To test this validity, we estimated nine paired-samples t-tests. As shown in Table 8, we found that for age and the manipulations of the influencer advertisements there were some differences in the relative attractiveness of the model and the product in the ads. Because age is to some extent related to attraction, some differences are to be expected, although when considering the effect sizes (mostly small to medium) the impact seems limited. Also, as shown in Table 9, for time spent we see a similar pattern, with no notable differences between the professional and AI-generated ads. Overall, the results support the internal validity of most AI-generated ad pairs.

To evaluate ecological validity, we compared the six AI-generated ads with the professional ad that served as the basis for the AI-generated ads on the three subdimensions of ecological validity. As shown in Table 9, we ran nine repeated-measured ANOVAs. First, the results indicate that, across all ads, the scores on quality, appropriateness, and realism are high. Of 18 AI-generated ads, 15 were considered equal in quality to the professionally created ads; for the dimensions of appropriateness and realism, the equal-in-quality number was 11. Despite differences between individual ads, there was a clear trend indicating that the professional ads generally scored comparably with the generative AI ads we developed. A systematic overview of the comparisons can be found in Table 10. Overall, these findings support the ecological validity of the AI-generated ad pairs.

Table 7. Results construct validity test.

Variable	Type of Ad	Conditions		t	p	Cohen's d
Age		Younger	Older			
	Influencer	26.4	44.3	32.7	$<.001$	1.64
	Car	28.9	56.7	45.6	$<.001$	2.28
	Fragrance	27.7	63.5	83.3	$<.001$	4.16
Expression		Smile	Neutral	t	p	Cohen's d
	Influencer	6.4	3.1	38.9	$<.001$	1.95
	Car	6.0	3.2	35.7	$<.001$	1.78
	Fragrance	4.7	2.9	27.6	$<.001$	1.38
Product		Product A	Product B			
	Influencer	99.8%	99.5%			
	Car	92.0%	99.5%			
	Fragrance	100%	99.5%			

Note. Degrees of freedom are 399.

Although external validity was not explicitly assessed, the fact that our empirical illustration showed that it was possible to implement our manipulations in different types of ads (i.e., social media influencer post, perfume ad, and car ad) without strong differences in terms of construct validity, internal validity and ecological validity can be seen as an initial indication that using generative AI for experimental stimuli creation can support the external validity of experimental advertising research.

Executing

Use in Study

An actual study in which the effects of our manipulations on certain cognitive affective or behavioral ad- or

Table 8. Results internal validity tests.

Variable	Type of Ad	Conditions		t	p	Cohen's d
Attractiveness model		Younger	Older			
	Influencer	6.1	5.6	−7.4	<.001	−0.37
	Car	6.1	4.9	−15.3	<.001	−0.76
	Fragrance	6.0	5.9	−2.6	.009	−0.13
		Smile	Neutral	t	p	Cohen's d
	Influencer	6.0	5.6	−7.9	<.001	−0.40
	Car	6.1	6.1	−0.9	.361	−0.05
	Fragrance	5.9	6.0	1.7	.093	0.08
Attractiveness product		Product A	Product B	t	p	Cohen's d
	Influencer	5.3	4.8	−6.4	<.001	−0.32
	Car	6.2	6.2	−1.1	.285	−0.05
	Fragrance	5.7	5.7	−0.2	.832	−0.01

Note. Degrees of freedom are 399.

Table 9. Comparison internal and ecological validity measures AI-generated and professional ads.

		Source						
		AI Generated						
Variable	Type of Ad	Younger	Older	Smile	Neutral	Product A	Product B	Professional
Time Spent[a] (Internal validity)	Influencer	41.2^{a}	44.8^{a}	42.4^{a}	40.5^{a}	40.8^{a}	43.6^{a}	50.0^{a}
	Car	35.7^{a}	36.6^{a}	33.0^{a}	38.1^{a}	37.4^{a}	37.4^{a}	39.2^{a}
	Fragrance	39.2^{ab}	43.6^{ab}	37.2^{ab}	36.7^{ab}	33.1^{a}	40.4^{ab}	38.6^{b}
Quality[b] (Ecological validity)	Influencer	5.3^{cde}	5.1^{abcd}	5.2^{bcde}	5.0^{abc}	5.3^{cde}	5.2^{abcde}	5.1^{abcd}
	Car	6.1^{cdef}	5.6^{a}	6.1^{def}	6.0^{bcd}	6.0^{bcd}	5.9^{bcd}	6.0^{bcd}
	Fragrance	5.9^{cdef}	5.8^{bcdef}	5.7^{ab}	5.8^{bcde}	5.8^{bcde}	5.6^{ab}	6.0^{def}
Appropriateness[b] (Ecological validity)	Influencer	5.9^{d}	5.7^{bc}	5.9^{d}	5.7^{abc}	5.9^{d}	5.5^{ab}	5.9^{d}
	Car	6.1^{bcd}	5.7^{a}	6.2^{de}	6.2^{cde}	6.2^{bcd}	6.1^{bc}	6.2^{dce}
	Fragrance	6.0^{bc}	6.0^{bc}	6.1^{bcd}	6.1^{bcd}	6.2^{cde}	4.8^{a}	6.3^{de}
Realism[b] (Ecological validity)	Influencer	5.3^{cd}	5.1^{abc}	5.2^{bcd}	4.9^{abcd}	5.3^{cd}	4.9^{ab}	5.1^{abc}
	Car	6.0^{bcd}	5.6^{a}	6.1^{cd}	6.1^{cd}	6.0^{bcd}	5.9^{bc}	6.0^{bcd}
	Fragrance	5.9^{b}	5.8^{b}	5.8^{b}	5.8^{b}	5.9^{b}	5.2^{a}	6.1^{c}

Note. Means, in the same row, that do not share a subscript$^{a–f}$ differ significantly ($p < .05$) from one another. Stated differently, means with a shared subscript$^{a–f}$ do not differ.
[a]Time spent is measured in seconds.
[b]Scales range from 1 to 7.

Table 10. Systematic overview differences internal and ecological validity measures AI-generated and professional ads.

		Manipulation								
Variable	Type of Ad	Younger vs. Older	Younger vs. Professional	Older vs. Professional	Smile vs. Neutral	Smile vs. Professional	Neutral vs. Professional	Product A vs. B	Product A vs. Professional	Product B vs. Professional
Time spent (Internal validity)	Influencer	N	N	N	N	N	N	N	N	N
	Car	N	N	N	N	N	N	N	N	N
	Fragrance	N	N	N	N	N	N	N	D	N
Quality (Ecological validity)	Influencer	N	N	N	N	N	N	N	N	N
	Car	D	N	D	N	N	N	N	N	N
	Fragrance	N	N	N	N	D	N	D	N	D
Appropriateness (Ecological validity)	Influencer	D	N	D	D	N	D	D	N	D
	Car	D	N	D	N	N	N	N	N	N
	Fragrance	N	D	D	N	N	N	D	N	D
Realism (Ecological validity)	Influencer	N	N	N	N	N	N	D	N	N
	Car	D	N	D	N	N	N	N	N	N
	Fragrance	N	D	D	N	D	D	D	D	D

Note. N = Not different; D = Different

brand-related consumer responses were tested was not part of our empirical illustration. Based on the pretest, it can be concluded that the items we used to assess the different forms of validity are useful to be implemented in actual advertising experiments.

Furthermore, when implementing AI-generated ads in an actual study it is also key to disclose the use of AI in a debriefing. At the end of our empirical study, we asked participants about the purpose of our study, and only a few (67 of 400 participants) included something related to AI in their answers. This finding can be seen as a pressing argument for the use of such disclosure.

Discussion

This paper provides a methodological contribution to advertising research by describing how generative AI can be used to create experimental stimuli. Our framework, the MADE (Mapping, Assembling, Demonstrating, Executing) framework, offers guidance and practical advice for researchers on how generative AI can be leveraged to enhance the quality and realism of stimuli. Our findings show that generative AI is an adept tool for creating stimuli, as evidenced by the fact that in more than two-thirds of cases AI could produce ads with quality as good as those professionally created by humans. As generative AI continues to improve, its potential to boost manipulation quality and the perceived realism of experimental stimuli in studies will grow, likely boosting different forms of validity including internal validity and most importantly ecological validity or realism in experimental advertising research. In short, our results show generative AI may help us conduct experiments that more accurately capture how consumers respond to advertising.

In the current paper, we demonstrated how the MADE framework can be used to develop visual stimuli for advertising research. However, the framework is designed to be versatile and adaptable for developing a wide range of stimuli beyond just visual ones. The framework's holistic approach allows it to be applied in different areas of advertising research, including textual, auditory, audiovisual, and potentially even interactive and immersive formats. Currently, established AI tools allow for the creation of textual ads, ad copy, slogans, and scenarios, and offer the ability to create auditory stimuli, such as radio advertisements and songs with brand placements. Emerging tools, such as OpenAI Sora and Google Veo, promise the creation of video content based on textual input. Moving forward, tools like these might enable the creation of video ads. It is likely, however, that the technological development will not stop with video content, and it is conceivable that, in the future, generative AI can be used to create stimuli in other ad formats too-such as interactive advertisements, advergames, and augmented reality/virtual reality advertising (Ahn, Kim, and Kim 2022; Terlutter and Capella 2013; Van Berlo et al. 2021; Van Berlo, Van Reijmersdal, and Waiguny 2023).

Whereas the advent of generative AI offers promising avenues for creating higher quality stimuli, our exploration also reveals an interesting paradox: Use of generative AI seems to increase the workload on researchers. Although the tools themselves are relatively quick at generating images, this action is only a small part of the overall process. The time increase arises from the need for the researchers to assess and use multiple tools, to iterate in refining prompts, to review numerous versions of generated content, to engage in additional editing, and potentially to conduct more extensive pretesting. So, contrary to the commonly held assumption that generative AI will streamline the process of creating stimuli, we surprisingly found that it can make stimuli creation more time-consuming for researchers (at least the first time).

Despite the time involved in using generative AI to create experimental stimuli, our results show the created stimuli can be of very high quality. The low cost and relatively low learning curve for generative AI thus can potentially elevate standards for stimuli. Elevating standards for research quality is ultimately beneficial for the field, ensuring that studies are more realistic and thus potentially more impactful. Despite the time investment currently needed to use generative AI, it is also likely researchers will become more efficient as they become more familiar with these tools. In addition, the tools themselves are likely to improve as well.

Based on our experiences with using generative AI for experimental stimuli creation both in the process toward formulating the MADE framework and for empirically illustrating the usefulness of this framework, we identified some observations, tips, and insights (see Table 11) that may help advertising researchers who are not yet familiar with using generative AI to get started with this approach.

Whereas the focus of this paper is on the use of generative AI to create experimental stimuli, our findings can also be extrapolated to infer that consumers seem to generally respond well to advertising created by generative AI. This response has potentially profound implications for the ad industry, particularly for those in ad creation. Our findings confirm that ads can be effectively created by generative AI, showing that an era of automated production of a greater variety of ad versions and personalization has begun.

Our study identifies several promising avenues for future research. A critical area is understanding consumer awareness and detection of generative AI-created images and how these factors influence subsequent

Table 11. Observations, tips, and insights on using generative AI for stimuli creation.

Creating

- **Prompt Engineering for Image Descriptions**: Utilize generative AI tools like ChatGPT or Midjourney's "/describe" function to craft detailed text prompts. This accelerates the process of generating suitable visual stimuli.
- **Tool Selection Based on Functionality**: Different generative AI platforms excel at various tasks, such as creation versus editing. Employ a combination of tools to achieve optimal results.
- **Iterative Prompt Refinement**: If initial outputs are unsatisfactory, refine the prompt and reattempt. This iterative process can significantly enhance output quality.
- **Leveraging Multiple Outputs**: Many AI tools can generate multiple versions of outcome from a single prompt. Use this option to increase efficiency and quality.
- **Navigating Brand Representation**: Given AI restrictions on generating branded content, incorporate brand elements (e.g., logos, product images) manually post-creation.
- **Session Continuity and Learning**: AI tools often adapt based on ongoing interactions meaning that creation can improve with continued use in the same session. If results are not as desired, starting a new thread may be beneficial.
- **Descriptive and Specific Prompts**: Detailed prompts, especially those specifying a desire for photorealistic results, generally yield better outputs.
- **(In)Consistency in Models:** When developing a pair of stimulus ads (with a single manipulation), know that not all AI tools enable the same model to be maintained between the different versions.

Assessing

- **Validity Checks**: Implement checks for manipulation, confound, and ecological validity to ensure the validity of AI-generated stimuli.
- **Realism**: For experiments a manipulation does not have to be photorealistic; perceptual pervasiveness and authenticity are often more important.
- **Ecological Versus Internal Validity**: Strive for a balance that does not compromise the ecological validity of the experiment while ensuring manipulation integrity.
- **Attention to Detail**: AI struggles with rendering realistic human features (e.g., faces, hands). Simplify requests concerning these details or plan for manual adjustments.

Using

- **Applications**: Generative AI-created ads can be used both in laboratory experiments and field experiments.
- **Context Integration**: Embed AI-generated visuals in realistic contexts (e.g., digital news platforms, social media feeds) to enhance ecological validity.
- **Ethical Considerations and Disclosure**: Ensure transparency about the use of generative AI with participants, either through prior disclosure or post-experiment debriefing, to respect ethical guidelines and participant informed consent. Remain ultimately responsible for scientific output.

General Recommendations

- **Multiple AI Tools**: Try out at least two generative AI tools when developing stimulus materials. There are often notable differences between different outputs.
- **Multiple Prompts**: Try at least two different versions of prompts since slight differences in phrases can often yield dramatic differences.
- **Multiple People**: Have at least two people try to create the images. This approach is a natural way to generate differences in which prompts are used and how they are tweaked during image creation.

consumer responses. As generative AI evolves, it raises questions about consumers' ability to distinguish AI-generated images from those created by humans, as well as the role that potential disclosures might play. Future research should take a similar approach to that of Arango, Singaraju, and Niininen (2023) to examine how well consumers can discern AI-generated ads and how this awareness affects key advertiser outcomes, such as ad engagement, trustworthiness, and, ultimately, persuasion.

Additionally, investigating the impact of generative AI across various consumer segments will provide valuable insights into the broader effectiveness and potential of AI imagery in advertising. Future research might explore how perceptions of AI-generated ads shift as consumers become more familiar with such content, given that some may be more adept at identifying AI-generated materials. Furthermore, attitudinal persuasion knowledge might serve as an important moderator for the effectiveness of AI-generated ads.

Finally, while this study demonstrates that generative AI can effectively create stimulus materials with varying levels of concrete ad characteristics (e.g., model age, product color), future research should employ the MADE framework to explore how generative AI can also be used to manipulate more abstract characteristics, such as creativity, vividness, or nostalgia. We believe that generative AI and the MADE approach have the potential to develop stimuli for these abstract characteristics, although the process may take longer and require more samples to be created and pretested. We encourage future research to explore this potential because it would provide a deeper understanding of the capabilities and applications of AI in advertising research.

This study, like all research, has its limitations. First, our research examined a limited selection of ads and manipulations, and there is certainly need for further studies that explore a wider range of advertising content and manipulative techniques. Furthermore, in the current study we used existing ads as starting points for our prompts, and future research should compare the quality of outputs produced using various types of prompts across different generative AI tools.

Additionally, the design of our study may introduce certain limitations. We selected a within-subjects design over a between-subjects design to leverage its greater power and sensitivity for detecting subtle differences across conditions. However, within-subjects designs inherently carry risks such as carryover effects, in which experiences in one condition might influence responses in subsequent conditions, and demand characteristics, in which exposure to all conditions might make participants more aware of the study's purpose. To mitigate these risks, we employed counterbalancing to ensure that the order of image pairs was randomized, making it impossible for participants to encounter them in consecutive order. Furthermore, participants were not permitted to revisit previous images, which helped limit any influence from prior conditions. However, we also suggest that future studies replicate our findings using a between-subjects design to confirm our results and potentially strengthen external validity.

Finally, even though we assessed a variety of validities in our study, statistical conclusion validity was not explicitly assessed. Future studies should address this validity by comparing the effects of manipulations using generative AI tools with traditional manipulations. Also, it should be noted that the rapid evolution of generative AI tools means that our findings, while current, may need to be revisited as these technologies progress.

To conclude, while the use of generative AI in advertising research presents methodological challenges, it also offers significant opportunities to enhance the quality and realism of experimental stimuli, and therefore represents a new chapter in solving the "realism–control paradox." As the advertising research field adapts to these technological changes, ongoing research is needed to further optimize the use of AI tools to create effective advertising research stimuli.

Acknowledgments

We want to thank Morgan Valent and Farhad Modarres Yazdi for helping us with the identification and testing of the AI tools.

Disclosure Statement

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

Notes on contributors

Zeph M. C. van Berlo (PhD, University of Amsterdam) is an assistant professor of persuasive communication, Amsterdam School of Communication Research (ASCoR), University of Amsterdam.

Colin Campbell (PhD, Simon Fraser University) is an associate professor of marketing, Knauss School of Business, University of San Diego.

Hilde A. M. Voorveld (PhD, University of Amsterdam) is an associate professor of persuasive communication, Amsterdam School of Communication Research (ASCoR), University of Amsterdam.

ORCID

Zeph M. C. van Berlo http://orcid.org/0000-0002-1008-8654
Colin Campbell http://orcid.org/0000-0002-6218-0866
Hilde A. M. Voorveld http://orcid.org/0000-0002-6667-3529

Data Availability Statement

The data that support the findings of this study, and the materials that we used, are openly available on OSF at https://osf.io/gzyr9.

References

Ahn, Sun Joo (Grace), Jooyoung Kim, and Jeamin Kim. 2022. "The Bifold Triadic Relationships Framework: A Theoretical Primer for Advertising Research in the Metaverse." *Journal of Advertising* 51 (5): 592–607. https://doi.org/10.1080/00913367.2022.2111729.

APA (American Psychological Association). 2023. "APA Journals Policy on Generative AI: Additional Guidance." American Psychological Association. https://www.apa.org. https://www.apa.org/pubs/journals/resources/publishing-tips/policy-generative-ai.

Arango, Luis, Stephen P. Singaraju, and Outi Niininen. 2023. "Consumer Responses to AI-Generated Charitable Giving Ads." *Journal of Advertising* 52 (4): 486–503. https://doi.org/10.1080/00913367.2023.2183285.

Audi. 2020. "Audi E-Tron Sportback [Advertisement]." https://www.audi-mediacenter.com/en/press-releases/re-orientation-of-the-brand-takes-shape-audi-starts-a-new-brand-campaign-13249.

Azrout, Rachid, Zeph M. C. Van Berlo, Roeland Dubèl, Jeroen G. Jonkman, Lara N. Wolfers, and Susanne E. Baumgartner. 2024. "Has the AI Revolution Started Yet? A Status Report on the Use of and The Attitudes Towards ChatGPT in Dutch Society." *Tijdschrift voor Communicatiewetenschap* 52 (3): 267–287. https://doi.org/10.5117/TCW2024.3.002.AZRO.

Banaji, Mahzarin R., and Robert G. Crowder. 1989. "The Bankruptcy of Everyday Memory." *American Psychologist* 44 (9): 1185–1193. https://doi.org/10.1037/0003-066X.44.9.1185.

Bockting, Claudi L., Eva A. M. Van Dis, Robert Van Rooij, Willem Zuidema, and Johan Bollen. 2023. "Living Guidelines for Generative AI—Why Scientists Must Oversee Its Use." *Nature* 622 (7984): 693–696. https://doi.org/10.1038/d41586-023-03266-1.

brihallofficial. 2018. "#Ad Bringing in the End of This Year. [Advertisement]." Instagram. https://www.instagram.com/p/BdQLozqA6WC/?next=/p/CizZ9BLuRDM/.

Campbell, Colin, Kirk Plangger, Sean Sands, and Jan Kietzmann. 2022a. "Preparing for an Era of Deepfakes and AI-Generated Ads: A Framework for Understanding Responses to Manipulated Advertising." *Journal of*

Advertising 51 (1): 22–38. https://doi.org/10.1080/00913367.2021.1909515.
Campbell, Colin, Kirk Plangger, Sean Sands, Jan Kietzmann, and Kenneth Bates. 2022b. "How Deepfakes and Artificial Intelligence Could Reshape the Advertising Industry: The Coming Reality of AI Fakes and Their Potential Impact on Consumer Behavior." *Journal of Advertising Research* 62 (3): 241–251. https://doi.org/10.2501/JAR-2022-017.
Campbell, Colin, Sean Sands, Carla Ferraro, Hsiu-Yuan Tsao, and Alexis Mavrommatis. 2020. "From Data to Action: How Marketers Can Leverage AI." *Business Horizons* 63 (2): 227–243. https://doi.org/10.1016/j.bushor.2019.12.002.
Chui, Michael, Eric Hazan, Roger Roberts, Alex Singla, Kate Smaje, Alex Sukharevsky, Lareina Yee, and Rodney Zemmel. 2023. "The Economic Potential of Generative AI: The Next Productivity Frontier." Bay Area, CA: McKinsey & Company. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier.
Cohen, I. Glenn, Theodoros Evgeniou, and Martin Husovec. 2023. "Navigating the New Risks and Regulatory Challenges of GenAI." Harvard Business Review. https://hbr.org/2023/11/navigating-the-new-risks-and-regulatory-challenges-of-genai.
Cook, Thomas D., and Donald T. Campbell. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin Harcourt.
Creswell, John W., and Cheryl N. Poth. 2016. *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. Thousand Oaks, CA: Sage publications.
De Pelsmacker, Patrick. 2021. "What Is Wrong with Advertising Research and How Can We Fix It?" *International Journal of Advertising* 40 (5): 835–848. https://doi.org/10.1080/02650487.2020.1827895.
Deveau, Richelle, Sonia J. Griffin, and Steve Reis. 2023. "AI-Powered Marketing and Sales Reach New Heights with Generative AI." Los Angeles, CA: McKinsey & Company. https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/ai-powered-marketing-and-sales-reach-new-heights-with-generative-ai.
Dom Qwek. n.d. "Dom Qwek." Accessed August 20, 2024. https://www.domqwek.com/artist-bio.
Eisinga, Rob, Manfred Grotenhuis, and Ben Pelzer. 2013. "The Reliability of a Two-Item Scale: Pearson, Cronbach, or Spearman-Brown?" *International Journal of Public Health* 58 (4): 637–642. https://doi.org/10.1007/s00038-012-0416-3.
Endert, Julius. 2024. "Generative AI Is the Ultimate Disinformation Amplifier." DW Akademie. https://p.dw.com/p/4doP4.
European Commission. 2024. "Living Guidelines on the Responsible Use of Generative AI in Research: An ERA Forum Stakeholders' Document." Publications Office of the European Union. https://research-and-innovation.ec.europa.eu/document/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en.
European IP Helpdesk. 2023. "Intellectual Property in ChatGPT." https://intellectual-property-helpdesk.ec.europa.eu/news-events/news/intellectual-property-chatgpt-2023-02-20_en.
Franke, Claudia, Andrea Groeppel-Klein, and Katrin Müller. 2023. "Consumers' Responses to Virtual Influencers as Advertising Endorsers: Novel and Effective or Uncanny and Deceiving?" *Journal of Advertising* 52 (4): 523–539. https://doi.org/10.1080/00913367.2022.2154721.
Geuens, Maggie, and Patrick De Pelsmacker. 2017. "Planning and Conducting Experimental Advertising Research and Questionnaire Design." *Journal of Advertising* 46 (1): 83–100. https://doi.org/10.1080/00913367.2016.1225233.
Grigsby, Jamie L., Robert D. Jewell, and César Zamudio. 2023. "A Picture's Worth a Thousand Words: Using Depicted Movement in Picture-Based Ads to Increase Narrative Transportation." *Journal of Advertising* 52 (4): 594–612. https://doi.org/10.1080/00913367.2022.2077267.
Hamby, Anne, and Cristel Russell. 2023. "Strategies and Tools to Enhance the Use of Dynamic Audiovisual Stimuli in Academic Advertising Research." *Journal of Advertising* 52 (4): 633–642. https://doi.org/10.1080/00913367.2022.2107121.
Hayes, Jameson L., Brian C. Britt, William Evans, Stephen W. Rush, Nathan A. Towery, and Alyssa C. Adamson. 2021. "Can Social Media Listening Platforms' Artificial Intelligence Be Trusted? Examining the Accuracy of Crimson Hexagon's (Now Brandwatch Consumer Research's) AI-Driven Analyses." *Journal of Advertising* 50 (1): 81–91. https://doi.org/10.1080/00913367.2020.1809576.
Hofer, Matthias. 2018. "Confoundation Check." In *The International Encyclopedia of Communication Research Methods*, edited by Jörg Matthes, Christine S. Davis, and Robert F. Potter. Hoboken, NJ: Wiley-Blackwell. https://doi.org/10.1002/9781118901731.iecrm0039.
Huh, Jisu, Michelle R. Nelson, and Cristel A. Russell. 2023. "ChatGPT, AI Advertising, and Advertising Research and Education." *Journal of Advertising* 52 (4): 477–482. https://doi.org/10.1080/00913367.2023.2227013.
Lehnert, Kevin, Brian D. Till, Kyle Coble, and Mark J. Arnold. 2023. "The Creative Knife: Does Interference Dull the Edge of Creativity Effectiveness?" *Journal of Advertising* 52 (4): 558–577. https://doi.org/10.1080/00913367.2022.2053900.
Louis Vuitton. 2018. "Les Perfumes [Advertisement]." https://uk.louisvuitton.com/images/is/image/lv/1/LV/louis-vuitton–LV_WF_Campaign_AR_Visual9.jpg.
McGrath, Joseph E. 1981. "Dilemmatics : The Study of Research Choices and Dilemmas." *American Behavioral Scientist* 25 (2): 179–210. https://doi.org/10.1177/000276428102500205.
Miao, Fengchun, and Wayne Holmes. 2023. *Guidance for Generative AI In Education and Research*. Paris, France: United Nations Educational, Scientific and Cultural Organization (UNESCO). https://doi.org/10.54675/EWZM9535.
Moorman, Christine, Harald J. Van Heerde, C. Page Moreau, and Robert W. Palmatier. 2019. "Challenging the Boundaries of Marketing." *Journal of Marketing* 83 (5): 1–4. https://doi.org/10.1177/0022242919867086.
Morales, Andrea C., On Amir, and Leonard Lee. 2017. "Keeping It Real in Experimental Research—Understanding When, Where, and How to Enhance Realism and Measure Consumer Behavior." *Journal of*

Consumer Research 44 (2): 465–476. https://doi.org/10.1093/jcr/ucx048.

Perdue, Barbara C., and John O. Summers. 1986. "Checking the Success of Manipulations in Marketing Experiments." *Journal of Marketing Research* 23 (4): 317–326. https://doi.org/10.2307/3151807.

Peres, Renana, Martin Schreier, David Schweidel, and Alina Sorescu. 2023. "On ChatGPT and Beyond: How Generative Artificial Intelligence May Affect Research, Teaching, and Practice." *International Journal of Research in Marketing* 40 (2): 269–275. https://doi.org/10.1016/j.ijresmar.2023.03.001.

Pot, Kim. 2023. "Beeldcreatie in DALL-E 3: Zo Schrijf je de Perfecte Prompt." Frankwatching. https://www.frankwatching.com/archive/2023/11/09/tips-prompt-dalle-3/.

Rivas, Pable, and Liang Zhao. 2023. "Marketing with ChatGPT: Navigating the Ethical Terrain of GPT-Based Chatbot Technology." *AI* 4 (2): 375–384. https://doi.org/10.3390/ai4020019.

Rodgers, Shelly. 2021. "Themed Issue Introduction: Promises and Perils of Artificial Intelligence and Advertising." *Journal of Advertising* 50 (1): 1–10. https://doi.org/10.1080/00913367.2020.1868233.

Sands, Sean, Colin Campbell, Kirk Plangger, and Carla Ferraro. 2022. "Unreal Influence: Leveraging AI in Influencer Marketing." *European Journal of Marketing* 56 (6): 1721–1747. https://doi.org/10.1108/EJM-12-2019-0949.

Schmidt, Frank L., and Jay E. Hunter. 2015. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Thousand Oaks, CA: Sage Publications. https://doi.org/10.4135/9781483398105.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin Harcourt.

Stam, Aleda. 2023. "AI Will Impact at Least Half of All Ad Revenue in 2023, GroupM Predicts." AdAge. https://adage.com/article/agency-news/ai-will-impact-least-half-all-ad-revenue-2023-groupm-predicts/2499161.

Terlutter, Ralf, and Michael L. Capella. 2013. "The Gamification of Advertising: Analysis and Research Directions of In-Game Advertising, Advergames, and Advertising in Social Network Games." *Journal of Advertising* 42 (2-3): 95–112. https://doi.org/10.1080/00913367.2013.774610.

Thomas, Veronica L., and Kendra Fowler. 2021. "Close Encounters of the AI Kind: Use of AI Influencers as Brand Endorsers." *Journal of Advertising* 50 (1): 11–25. https://doi.org/10.1080/00913367.2020.1810595.

U.S. Congress. House. 2023. "AI Disclosure Act of 2023." HR 3831, 118th Cong. Adopted May 6, 2023. https://www.congress.gov/bill/118th-congress/house-bill/3831.

Vakratsas, Demetrios, and Xin (Shane) Wang. 2021. "Artificial Intelligence in Advertising Creativity." *Journal of Advertising* 50 (1): 39–51. https://doi.org/10.1080/00913367.2020.1843090.

Van Berlo, Zeph M. C., Eva A. Van Reijmersdal, Edith G. Smit, and L. Nynke Van der Laan. 2021. "Brands in Virtual Reality Games: Affective Processes Within Computer-Mediated Consumer Experiences." *Journal of Business Research* 122: 458–465. https://doi.org/10.1016/j.jbusres.2020.09.006.

Van Berlo, Zeph M. C., Eva A. Van Reijmersdal, and Martin K. J. Waiguny. 2023. "Twenty Years of Research on Gamified Advertising: A Systematic Overview of Theories and Variables." *International Journal of Advertising* 42 (1): 171–180. https://doi.org/10.1080/02650487.2022.2143098.

Van Esch, Patrick, Yuanyuan (Gina) Cui, and Shailendra P. Jain. 2021. "Stimulating or Intimidating: The Effect of AI-Enabled In-Store Communication on Consumer Patronage Likelihood." *Journal of Advertising* 50 (1): 63–80. https://doi.org/10.1080/00913367.2020.1832939.

Van Heerde, Harald J., Christine Moorman, C. Page Moreau, and Robert W. Palmatier. 2021. "Reality Check: Infusing Ecological Value into Academic Marketing Research." *Journal of Marketing* 85 (2): 1–13. https://doi.org/10.1177/0022242921992383.

Voorveld, Hilde A. M., Corine S. Meppelink, and Sophie C. Boerman. 2023. "Consumers' Persuasion Knowledge of Algorithms in Social Media Advertising: Identifying Consumer Groups Based on Awareness, Appropriateness, and Coping Ability." *International Journal of Advertising* 43 (6): 960–986. https://doi.org/10.1080/02650487.2023.2264045.

Wang, Chenyue, Sophie C. Boerman, Anne C. Kroon, Judith Möller, and Claes H. De Vreese. 2024. "The Artificial Intelligence Divide: Who Is the Most Vulnerable?" *New Media & Society*. https://doi.org/10.1177/14614448241232345.

Wu, Linwan, Naa A. Dodoo, and Taylor J. Wen. 2024. "Disclosing AI's Involvement in Advertising to Consumers: A Task-Dependent Perspective." *Journal of Advertising* 1–19. https://doi.org/10.1080/00913367.2024.2309929.

Yin, Robert K. 1994. "Discovering the Future of the Case Study Method in Evaluation Research." *Evaluation Practice* 15 (3): 283–290. https://doi.org/10.1016/0886-1633(94)90023-X.

Yoon, Julia, and Chris Borges. 2024. "Informing the Innovation Policy Debate: Key Concepts in Copyright Laws for Generative AI." Center for Strategic & International Studies. https://www.csis.org/blogs/perspectives-innovation/informing-innovation-policy-debate-key-concepts-copyright-laws.