Automated Alignment: Guiding Visual Generative AI for Brand Building and Customer Engagement

Tijmen P. J. Jansen
University of Hamburg
tijmen.jansen@uni-hamburg.de

Mark Heitmann
University of Hamburg
mark.heitmann@uni-hamburg.de

Martin Reisenbichler
University of Hamburg
martin.reisenbichler@uni-hamburg.de

David A. Schweidel
Emory University
dschweidel@emory.edu

Automated Alignment:

Guiding Visual Generative AI for

Brand Building and Customer Engagement

Abstract

Generative AI is poised to transform the way in which brands market to consumers. Recent research has demonstrated the potential benefits of AI in producing text, but limited work has examined how marketers can leverage AI to create visual assets. Visual elements play a crucial role in communicating with and engaging consumers, and visual generative AI has shown impressive performance in generating objective image content. But is it possible to train generative AI directly on marketing objectives? Making use of open-source generative AI tools, we propose a flexible algorithm that is informed by consumer responses to create images for marketing communications that are designed to achieve particular objectives. We illustrate how marketers can use their own visual assets and those of their competitors to tailor a generative AI for their brand's use. Our results show that producing content in this way can be more effective at various stages of the purchase funnel than the brand's own marketing content. Beyond funnel metrics, we demonstrate that generative AI can be used to convey specific dimensions of brand personality without compromising on brand engagement. Taken together, we discuss the implications for brand marketers and their agencies.

Keywords: Generative AI, Digital Marketing, Purchase Funnel, Brand Personality, AI Alignment

Introduction

Visual content is a critical element of a brand's marketing strategies. Research has probed the effects of different visual elements in marketing content on consumer outcomes such as attention and attitude toward the brand (Dall'Olio and Vakratsas 2023; Hartmann et al. 2021; Li and Xie 2020; Pieters and Wedel 2004; Pieters et al. 2010). Visual online advertising (i.e., display ads) has become a mainstay of digital marketing for driving sales and increasing market share, as well as cultivating a brand's desired image (Affonso and Janiszewski 2023; Brasel and Gips 2008; Dew et al. 2022; Pamuksuz et al. 2021), with brand investments in online display ads expected to exceed \$300 billion (Cramer-Flood 2023).

Against this backdrop, there is growing interest in understanding the capabilities of generative AI to support the creation of visual marketing content. While research has probed the applications of text-based generative AI (Brand et al. 2023; Reisenbichler et al. 2022), limited research has examined the capabilities of generative AI to support the creation of visual marketing content (Dew et al. 2022). Despite this, there are clear indications of brands' desires to leverage this emerging technology. For example, Coca-Cola invited digital artists to leverage AI to "Create Real Magic" by creating work to be featured on digital billboards in high-profile locations.[1] Relatedly, Heinz Ketchup has prompted visual generative AI to create various versions of ketchup bottle images, which they found to resemble the iconic Heinz bottle shape despite "Heinz" not being prompted.[2] Amazon has recently unveiled AI image generation to support brands that are advertising on its platform by facilitating generation based on product descriptions, making these editable by advertisers through text prompts.[3]

As these examples illustrate, firms are experimenting with different ways in which they can leverage generative AI. However, these campaigns rely on human intuition and creativity

[1] https://www.coca-colacompany.com/news/coca-cola-invites-digital-artists-to-create-real-magic-using-new-ai-platform
[2] https://campaignsoftheworld.com/digital/heinz-a-i-ketchup/
[3] https://advertising.amazon.com/blog/ai-image-generation

to guide the image creation. Humans prompt, whether directly or through textual product descriptions, what they would like to put on an image. Often, experimentation and editing are involved to yield impressive results. Open AI's DALL-E 3 has taken AI-supported image creation a step further by automatically refining the entered text prompt, effectively engaging in prompt engineering, to yield more desirable results.[4]

A critical limitation of the aforementioned examples is that they do not attempt to train the generative AI on outcomes relevant to marketing, such as sales or interest in the featured brand. Rather, to achieve such objectives, the intervention of a human would be necessary to provide the necessary guidance and align the AI output with the business objective. But, is it possible to turn machines into creative partners with a specific business objective in mind? That is, can machines go the full distance and generate entire advertising images that accomplish a brand's objectives without human content prompts or post hoc image editing?

The creation of visual advertisements to support a brand's marketing can be complex. The design of visual ads, for example, may require making trade-offs among different business objectives (Brasel and Gips 2008; Dall'Olio and Vakratsas 2023). The visual content that successfully encourages consumers to progress through the different stages of the purchase funnel (attention, interest, desire, and action; AIDA) (Batra and Keller 2016; Wedel and Pieters 2000) may be different from the visual content that effectively positions the brand along a particular brand personality dimension in the mind of the consumer (Keller and Lehmann 2006; Malär et al. 2011). Given the varied performance of different visual elements throughout the customer journey, brands may require multiple pieces of visual content, further adding to the expenses associated with content creation.

The advent of large-scale AI models for generating convincing, high-quality images like OpenAI's DALL-E, Midjourney (Oppenlaender 2022), Stable Diffusion (Rombach et al. 2022), or Adobe Firefly suggest that the costs associated with producing visual marketing

[4]https://openai.com/dall-e-3

content could suddenly fall. Despite their evident merits, these tools are designed to turn text prompts into objective image content, such as ketchup bottles in different contexts. Realizing the benefits of visual generative AI in marketing hinges on alignment between the generative AI process and the marketing objectives. To the best of our knowledge, research to date has not empirically demonstrated an efficient means of accomplishing this in the context of producing effective visual marketing content. We propose and demonstrate how a visual generative AI can be "fine-tuned" with a particular business objective.

In this research, we train a visual generative AI directly on mindset metrics (Colicev et al. 2018). Traditional ad generation often involves an iterative process among marketers, ad agencies and market researchers. If mindset metrics indicate ads do not perform as intended, agencies go back to the drawing board and attempt to better match brand KPIs. We endeavor to streamline this process by training a generative AI to produce content that is in line with a brand's desired positioning, or to engage consumers in terms of purchase funnel KPIs. Doing so would yield cost savings, as well as time savings, in the process of creating marketing content. However, this task is non-trivial. For example, researchers have questioned whether visual generative AI models are capable of learning (Ruiz et al. 2023) and generating abstract visual features that effectively engage consumers. In terms of ad effectiveness, it is therefore an empirical question as to how such content would compare to creatives developed through a traditional process.

We present an AI workflow to conceptualize and implement an integrated system for generating visual marketing content. Marketing input that needs to be provided by a user is limited to defining strategic brand objectives and selecting from the image output. We demonstrate the content creation algorithm by producing the visual elements to be used in digital display ads. We investigate the workflow's ability to produce advertisements that are effective at different stages of the purchase funnel, as well as that evoke specific dimensions of brand personality in the minds of consumers.

Across four empirical studies, our analyses reveal that the AI-generated ads can achieve superior performance on traditional funnel metrics that are typically collected and used as KPIs when advertising is being pre-tested (Smith et al. 2008). Our analyses also find that ads produced by the workflow outperform ads generated by prompting objective image content. Moreover, we find that the content generated by our algorithm evokes targeted brand personality dimensions (Aaker 1997; Dew et al. 2022; Dzyabura and Peres 2021) without undermining purchase funnel objectives. Taken together, our research shows the potential for generative AI to be a powerful partner in the creative process.

The remainder of this research is structured as follows. First, we briefly review the literature related to brand personality and visual content in marketing. We then describe the proposed algorithm, implemented using readily available open-source software, that generates new visual content based on the specified marketing objective. Finally, we present our empirical studies that apply our AI workflow in the context of the automotive industry. We conclude with a discussion of the implications of this research for both marketing researchers and practitioners.

Related Literature

Branding and Short-term Communication Objectives

Prior research on brand management depicts branding as being central to market success and, as such, as a top management priority in today's companies (Chaudhuri and Holbrook 2001; Keller and Lehmann 2006; Lovett et al. 2013; Pogacar et al. 2021). The literature is rich with illustrations of the complexities involved in branding, arising from cognitive, affective, and behavioral effects among consumers under various conditions (Chaudhuri and Holbrook 2001; Costello et al. 2023; Datta et al. 2017; Malär et al. 2011; Park et al. 2010).

Advertising and visual communication is a major channel through which brand equity is built. Visual communications seek to convey meaning on behalf of the brand, and in service of fulfilling specific goals such as driving sales, through the use of semantic elements, as well as specific visual elements such as colors and shapes (Affonso and Janiszewski 2023).

Advertising research has specified hierarchy of effects models (Smith et al. 2008) to indicate important advertisement success factors and typical processing hierarchies among consumers throughout the purchase funnel. The AIDA (attention, interest, desire, action) model is among the most cited, and arguably one of the most prominent models in the industry. Since it is challenging to isolate the impact of individual visual elements on the brand's bottom line, marketers conduct A/B survey tests using AIDA mindset metrics to understand the potential of new advertisements, choosing the most promising one and deciding whether to invest in ad distribution. Recent years have seen the application of machine learning methods to further assess how consumers react to visual brand communications (Dzyabura and Peres 2021; Liu et al. 2020), e.g., building on direct consumer responses to visual content (Hartmann et al. 2021).

Branding and Brand Identity

However, AIDA models do not tell the full story of advertising success. While attracting consumers is a relevant short-term objective, brands thrive in the long term by achieving an attractive and unique positioning in the hearts and minds of consumers. Without such brand positioning, brands would risk becoming exchangeable. This is of particular concern, should competing brands train comparable generative AI tools on similar funnel metrics.

Traditionally, brands differentiate by defining and coherently communicating a desired brand identity with the objective of building a unique brand image in the hearts and minds of consumers. Aaker (1997) identifies brand personality dimensions including "sincerity," "excitement," "competence," and "ruggedness" to measure brand image. Success is often measured in terms of perceived personality traits and whether ads succeed in evoking desired associations.

Malär et al. (2011) find that matching brand personality traits with the consumers' self-perception leads to more emotional attachment of consumers toward a brand. In a similar spirit, Liu et al. (2020) examines the personalities expressed by brands and consumers based on the images posted to social media in which the focal brand has been tagged, often finding a reasonable degree of alignment between the brand personalities inferred from the brands' own images and those posted by consumers. In light of the importance of brand personality, recent research continues to apply it to both, understand how consumers perceive a given brand (Dzyabura and Peres 2021), and to predict how the visual content used by brands will be perceived by consumers (Dew et al. 2022).

For these reasons, generalized models of advertising success factors in current research typically include AIDA-type properties, while also incorporating brand-related personality traits to ensure long-term brand building and differentiation (Batra and Keller 2016; Keller and Lehmann 2006). However, combining distinct visual elements to achieve these different objectives is a delicate task. Crafting advertising that effectively attracts customers to the brand in the short term may come at the expense of the brand's desired positioning. It remains an empirical question that we tackle in the present research as to how effectively generative AI can be used to produce visual content that addresses more complex marketing goals and mixes different brand objectives.

Designing Visual Advertising

Marketing research has also sought to identify guiding principles of how to design visuals such as logos and products, both in terms of a designer's working scheme and in terms of consumer perceptions. We briefly discuss two illustrative streams of literature that contribute to the development of our proposed AI workflow, which is intended to imitate the workflows of ad design practitioners (Reisenbichler et al. 2022).

Categorization theory of ideation

According to Dew et al. (2022), designers implicitly rely on the categorization theory of ideation when composing a new visual artifact. Visual ads and brands exist in a landscape of brand visualizations that form the basis of consumers' mental categories. Consumers evaluate new designs based on the concepts that those designs activate in their minds, and designers position their visual ads within the visual landscape. Consistent with the categorization theory of ideation, Dzyabura and Peres (2021) detect and illustrate typical visual representations of brands in consumers' minds.

Based on these insights, a generative AI must produce a sense of the general visual ad landscape in the industry of interest *and* incorporate favorable properties such as evoking attention and interest. As we will discuss, this can be achieved by training a large generative model with appropriate content.

Familiarity, fluency and beauty-in-averageness

In terms of visual content, proponents of fluency theory and beauty-in-averageness (Langlois and Roggman 1990) argue that more familiar content is easier and more fluent to process. This is an inherently positive experience that translates to liking and overall favorable responses (Reber et al. 2004). Toubia and Netzer (2017) find that familiarity is an important driver of successful innovations. Relatedly, Heitmann et al. (2020) illustrate how brands can position themselves by building on visual similarity to their competitors. As such, proximity to the visual mental design-prototype is key to success.

Training a generative AI on visual content from a specific context (e.g., a particular industry), may enable it to learn high-level averages that are associated with prototypical advertisements. Similarly, training a generative AI on visual content associated with particular brand personality dimensions may make it learn those elements that are more commonly associated with that dimension. This is similar to the way in which Liu et al. (2020) characterize brand- and user-generated content as evoking a particular brand personality. Generative AI might be able to directly incorporate these different marketing objectives into visual communication.

Using Generative AI to Create Visual Marketing Content

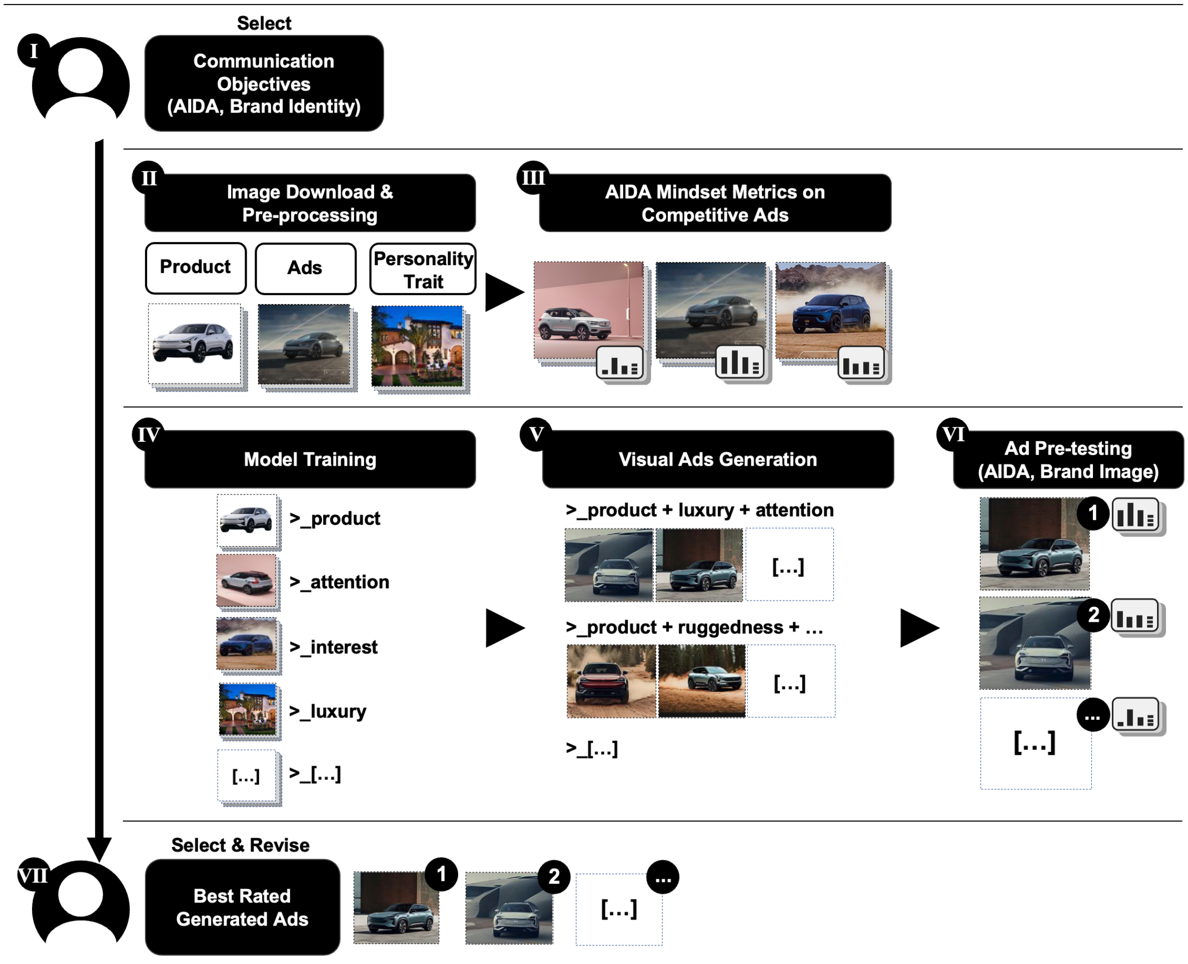
To demonstrate the potential for generative AI to produce high-performing visuals, we focus on the online display market for the automotive industry as an illustrative example. We select the Polestar 3 electric vehicle as an example product and brand. Due to its complexity, the task of generating ad visuals holds a number of significant methodological challenges. These include the need for managerial guidance to attain desired marketing objectives (Peukert et al. 2023; Proserpio et al. 2020), for integrating visual consumer preferences (Dzyabura and Peres 2021), for translating verbal concepts into suitable visual representations in ad creation (Rombach et al. 2022; Ruiz et al. 2023), as well as balancing multiple marketing objectives (Dew et al. 2022), and feedback loops (Kulczynski and Hook 2023).

We resolve these challenges by proposing an adaptive generative system, tailored for the problem of visual ad generation (see Figure 1). Our focus is on generating the visuals used in advertising. Other technologies can be used for text generation to complement our output (e.g., Reisenbichler et al. 2022). We aim to create intriguing, novel visuals that adapt to any perceptual marketing objective without the need for human ad creation. Below, we lay out details on typical challenges and our tailored solution.

I: Managerial Guidance of Ad Creation

In traditional agency settings, ad creation starts with a briefing of marketing on the desired communication objective(s). According to Keller (1993), building brand equity in the minds of consumers involves building brand image associations and brand awareness (purchase funnel KPIs). Accordingly, marketers specify the desired brand identity as well as additional short-term communication KPIs such as brand attention or interest. These two types of objectives need to be infused into generative AI pulling through all stages of ad creation (I, Figure 1). Similar to traditional ad creation, final selection of the output involves human evaluation informed by ad pre-testing (VI, Figure 1). Note, however, unlike traditional ad generation, generative AI can produce an unlimited amount of potential candidates, so pre-testing can include many times more alternatives than traditional A/B testing of a limited number of alternatives.

Figure 1: An Integrated Approach to Generating Visual Advertisements



II and III: Data Preparation - Integrating Consumer Preferences and Perceptions

Ad agencies aim at crafting ads that reflect visual consumer preferences and associations with visual elements (Dzyabura and Peres 2021). Similar in idea to Reisenbichler et al. (2022), an automated AI procedure needs to learn on top-performing content to incorporate factors that are associated with success when it comes to visually evoking brand personality traits or purchase funnel goals like attention. To achieve this, an automated procedure needs to contain a market research-like module that involves rating existing visual ads on marketing objectives to let the model learn on top-rated visual templates. To accomplish this, in our second step, we scrape and pre-process template images (II, Figure 1), and collect consumer ratings on many advertisements (III, Figure 1). This enables the generative AI to learn optimal visual cues for reaching abstract advertisement goals in consumers' minds.

More precisely, for II, we engage in gathering and pre-processing necessary data like image templates that depict the product, as well as typical car ads, and pictures associated with specific brand personality traits. This provides the source material from which the generative AI learns the visual style of successful ads, taking into account a specific marketing objective.

First, our procedure must be able to depict the brand's focal product ads are run for. That is why we scrape Polestar 3 product pictures from the Internet.[5]

Second, beyond learning the product that will be featured in the generated ads, the model must also learn the visual ad language of the product category (cars), including typical representations like color palettes, camera angles, and the surrounding environment. To steer the model toward generating images that follow conventional advertising language, we scraped 211,429 online ads from Google's ad systems using the web archive. Since we are interested in a particular product category (cars), we use the YOLO object detection algorithm to identify ads that include these products and further eliminate ads for other product categories. This leaves us with 543 unique car ads (see Appendix A for details on the full process).

Third, the final set of images that are necessary to support our proposed workflow will enable the generation of images that evoke specific dimensions of brand personality. While the previously collected images provide a sense for the visual advertising language for the category, these may not contain sufficient data for training unique brand identities. Following Liu et al. (2020), we scrape 1,000 Flickr images for the keywords of the desired personality traits. Given the positioning of the Polestar 3 electric vehicle, we collect images that are associated with "ruggedness" and "luxury".

Since images scraped from the Internet come in different sizes and forms, we take several

[5]Brands likely have the digital assets at hand and can simply use their internal resources.

pre-processing steps to standardize the model input, as is typical in computer vision tasks (Hartmann et al. 2021; Liu et al. 2020). Since our focus is on advertising visuals, these pre-processing steps include erasing text from the template ads using Keras-OCR for text detection and CV2 for inpainting to produce clean ad images. Lastly, images are resized to 512×512 pixels. For the product pictures, we automatically detect and delete the background so that we are left with only the product in the image and no surroundings. We erase the surroundings from the product picture to prevent any inadvertent interference with model training (e.g., the AI inadvertently forming a strong association between the background and the product).

Regarding III, similar to Dew et al. (2022), we infuse consumer-based perceptions of ads into our generative system. For this purpose, we recruit study participants on Prolific to provide human ratings of the scraped ads. Using a randomly assigned online survey, we collect metrics associated with different stages of the purchase funnel. Note that unlike traditional ad pre-testing on a few ad alternatives, we need to test a high number of ads to train the visual language to the generative AI, which limits the number of ratings that we can collect for each individual ad. We found that a minimum of 5 ratings per advertising image suffices to train the generative AI to produce images that perform well on independent ratings (see Ceylan et al. 2023; Peng et al. 2020; Troncoso and Luo 2022, for similar sample sizes when training and evaluating machine learning). Further details on the survey are shown in Appendix B.

In sum, these image gathering, and pre-processing, and survey-based selection steps lead to a set of product pictures, a set of highly rated ads according to AIDA dimensions, and a set of pictures associated with brand personality traits, yielding a set of fine-tuning samples for our generative model.

IV: Model Training - Merging Verbal and Visual Information

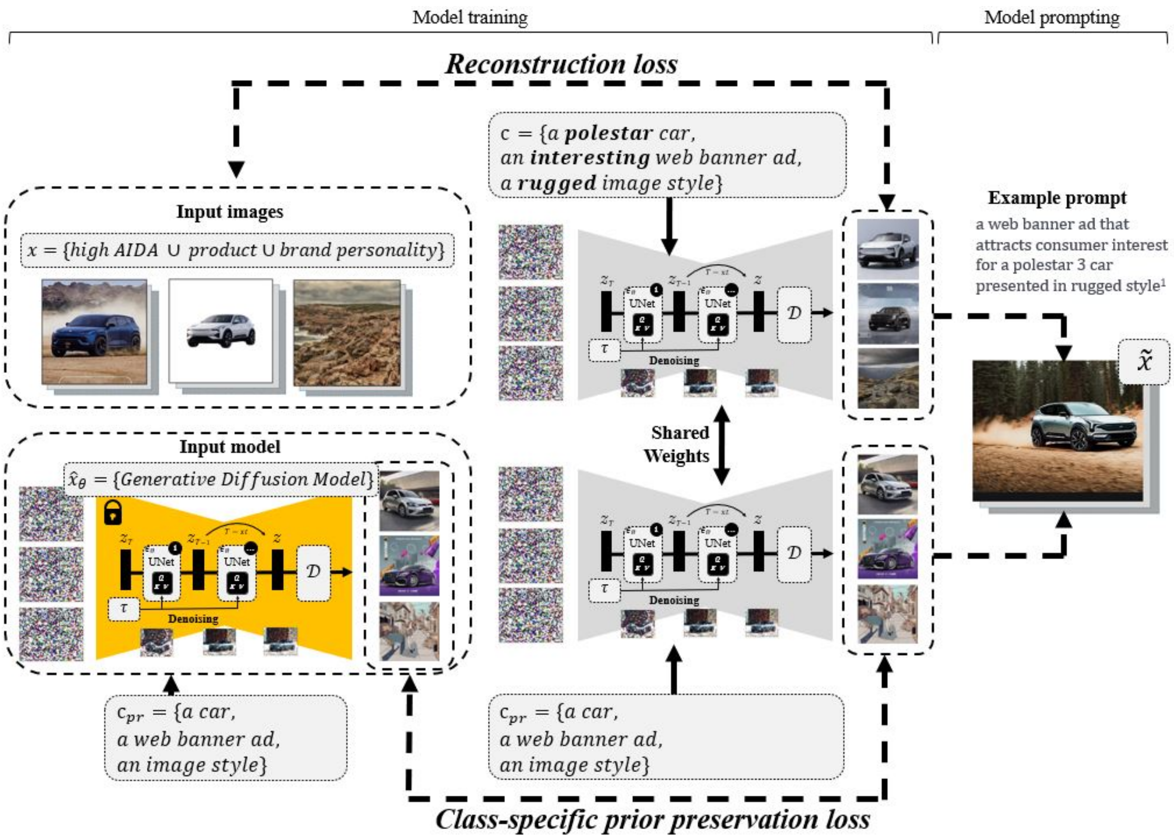
Designers in ad agencies merge various verbal and visual cues into ad drafts. Specifically, verbal marketing objectives need to be translated into visual communication. This can involve input from consumers such as collages (Zaltman and Coulter 1995) or verbal association maps (John et al. 2006), creative discussions based on mood-boards and (implicit) knowledge of typical representations in the target market.

Building on a good understanding of generally appealing visuals, designers can translate their understanding of the relevant visual language into advertising concepts. Similarly, we build on a foundational model that contains an understanding of generally appealing visual language that is flexible enough for feeding it with samples of various multi-modal (text and image based) concepts for model training. Specifically, we integrate verbal and visual knowledge (IV, Figure 1) by fine-tuning a pre-trained foundational generative AI model on desired marketing objectives and the visual language of ads in the product category as well the product itself (Figure 2). We feed the AI with example product, high performing ad, and brand personality pictures (x) and prompts (c) to steer the generated ads into specific marketing directions, while relying on concepts and prompts known to the pre-trained model, $\hat{x}_\theta$, like "car", "web banner ads", and "image styles" (Rombach et al. 2022; Ruiz et al. 2023).

To demonstrate our proposed workflow, we use the pre-trained generative diffusion model Stable Diffusion as a foundational model that we subsequently fine-tune. Generative diffusion models (GDMs) such as Stable Diffusion are set up as text-to-image generators, where text prompts like "*Generate a pink car*" can be used to generate a high-quality image representing this objective input (Rombach et al. 2022). Similar to the white canvas in a designer's mind with steps of fleshing out the idea when imagining a novel visual ad, *GDMs* generate images by removing noise in successive reverse diffusion (denoising) steps from a noisy image (Figure 2).

Formally speaking, the primary objective of GDMs like Stable Diffusion is to learn a data distribution $p(z)$, of a latent space representation (z) of an encoded original input image

Figure 2: Training Generative AI



[1]Note: Brand and product name, marketing KPI and brand personality dimension replaced by random letter combination to reduce potential language drift.

(x). This is achieved through a specific type of Convolutional Neural Networks (UNets) in sequential time steps, $t = 1...T$, which denoises a normally distributed initial noise map $\epsilon \sim \mathcal{N}(0, 1)$. After denoising from the latent space, z_T, z can be decoded back ($\mathcal{D}$) into a visual image or pixel space, effectively generating an output picture, $\tilde{x}$, (Rombach et al. 2022).

To gain better control over the model's output, the model is conditioned on text inputs (prompts, c) using a transformer model like CLIP τ (Ruiz et al. 2023). Transformers learn token-based embeddings of natural language to infuse prompts into the model for guiding image generation. The text prompts are integrated into the model through pre-trained attention weight matrices, which guide the denoising steps into specific visual directions as indicated by the text prompts.

Despite generic diffusion models' ability to be directed using prompts, it is not trained on perceptual marketing communication KPIs or brand image perceptions, as these vary by product category and target group. Suppose the desired marketing objective is to drive brand interest while conveying a rugged brand image. Simply prompting the standard model with such objectives does not achieve the desired results. To address this, we need to fine-tune the model with novel prompt and image pairs. Our goal in this step is to tune the generative model in the visual language of a specific marketing objective. For example, to create ads that generate high interest, we use ads with the highest interest-ratings for the fine-tuning process. In this way, the generative AI learns the visual language associated with high-interest in ads. The same approach can be used with any desired metric marketers are interested in, such as purchase intention or particular brand image dimensions, such as ruggedness.

We achieve this by tuning the denoising convolutional UNet model on the best performing image templates ($x = \{high\, AIDA \cup product \cup brand\, personality\}$, where *"high AIDA"* means top consumer rated ads along AIDA dimensions like interest, *"product"* denotes product pictures, and *"brand personality"* means pictures referring to brand personality traits like ruggedness. We thereby integrate a custom set of prompts, c, that relate to dimensions in x, and are not part of the GDM's prompt repertoire (Figure 2). The major issue with training new image content is language drift. Specifically, the GDM may effectively forget the general meaning of "car" when we train it on new vehicles. This is of particular concern for new abstract concepts like brand personality or purchase funnel objectives that might link to multiple existing ones in the GDM. Since we train the model on a limited number of examples, we might also reduce output variance and obtain lower quality results as a consequence.

To address these issues, we build on (Ruiz et al. 2023) and employ two loss functions, the *"Reconstruction loss"* and the *"Class specific prior preservation loss"* (see Figure 2). While the former ensures the model is properly tuning on the new concepts and input images, the latter prevents language drift and maintains the integrity of the latent space of Stable Diffusion by keeping the meaning and associated visual patterns of categories like "*car*", where the subscript *pr* denotes priors (i.e., model knowledge before tuning by us took place). This enables drawing from Stable Diffusion's pre-trained broad and impressive picture generation capability while enriching it with image templates related to marketing objectives and subjective consumer responses to generate an output image ($\tilde{x}$). It effectively links verbally expressed marketing concepts to the generic visual language of Stable Diffusion. This is similar in spirit to traditional briefings of creative agencies with the specific objectives and brand identity relevant for an individual campaign.

V: Ad Generation - Balancing Creativity and Constraints

Though it may seem to be a small component of the proposed workflow, the role of consumer input is crucial. First, ratings from consumers are used to identify the most relevant ads, which are then used to train the generative AI. The existing ads used to imbue the model with an understanding of the visual language of ads via training are chosen based on consumer perceptions on that particular dimension. This process ensures that the generated visuals are aligned with the marketer's objective. Those ads that yield the highest reported purchase intention, for example, employ different visual elements compared to those ads that elicit the highest reported interest. By collecting consumer perceptions of the existing ads on multiple dimensions, marketers have the flexibility to inform the model for any objective they may have. This same approach can be used for targeting ads at different customer segments with different interests and preferences. In our illustration, we focus on the market average.

While creativity in finding novel visualizations is a key success ingredient in ad crafting practice, ad managers and designers usually constrain it to the level where ad objectives like brand positioning are still fulfilled (Dew et al. 2022; Dzyabura and Peres 2021). Consequently, a generative AI needs to be flexible and master creative solutions, coming up with novel, surprising, and intriguing visuals to communicate abstract properties like brand personalities (e.g., rocky outdoor environments, or misty mountain ranges for the personality trait ruggedness), while staying true to specified KPI objectives. We steer generative AI by a combination of keeping basic concepts like "car", "web banner ads", and "visual styles" in the model, using our new training concepts, and via hyperparameter optimization and testing.[6] That way, we generate creative ads (V, Figure 1) that resemble consumer preferences and designers' concepts via rich visual clues - keeping the model's output creative and open, while working within the target constraints as specified in step I.

The purpose of our investigation is to test the ability of generative AI to select and produce image content geared at marketing objectives. This is different from prompting objective image content, as the updated GDM decides which combination of visual elements is best suited to attain the specified objective. We deliberately did not add any additional prompt components relating to additional objective image content. Doing so would be akin to prompt engineering, which would add a layer of human creativity and intuition, and would limit our ability to assess the capabilities of the AI workflow. We did, however, experiment with different orders of training concepts. Irrespective of the marketing KPI we studied, we found the best performing order of concepts in prompting the model to generate visual ads to be: $c = [\textit{rated ads} \cup \textit{product} \cup \textit{brand personality trait}]$.

VI: Feedback Integration - Ad Pre-Testing

Agencies and marketing managers usually rely on feedback cycles over their ad production phase to ensure branding and campaign KPIs will be met. A well-established form of integrating feedback is ad pre-testing based on consumer surveys (Kulczynski and Hook 2023). Our workflow is directly trained on consumer input and contains no human interpretation. This might suggest more reliable results and more consistent marketing performance.

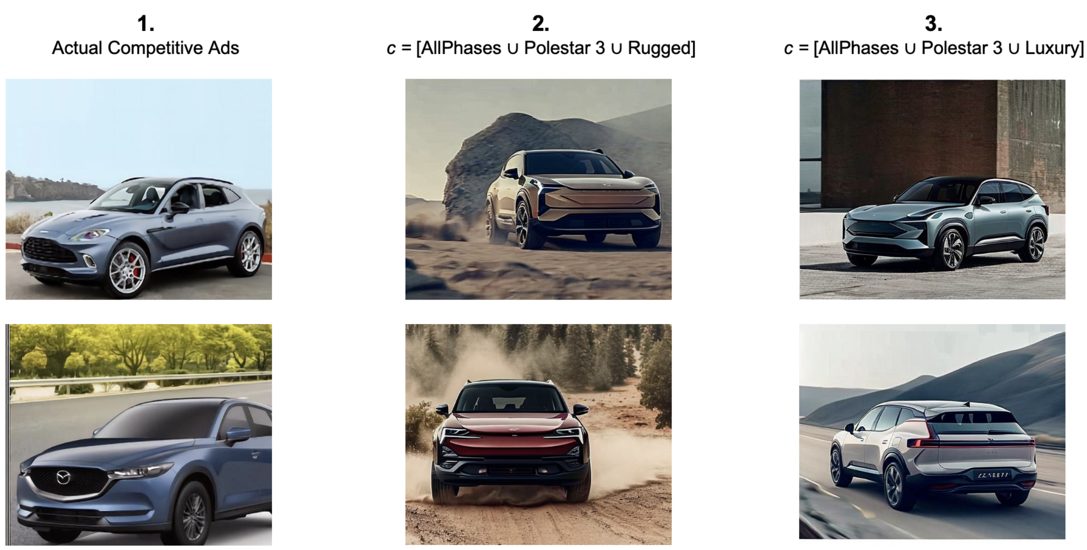
[6] This process requires extensive hyperparameter optimization for desirable results, including adjusting the guidance scale for the level of creativity vs. prompt orientation, and specifying inference steps for image denoising. After thorough testing, we set the guidance scale to 7.5 and the inference steps to 100. (Diab et al. 2022)

However, automated image generation is a probabilistic process that also results in varying quality for a chosen objective. Like Reisenbichler et al. (2022), we therefore rate our generative ad output to assess the ads' suitability for the brand. Using the generated ads, we use the same process of sourcing participants from Prolific to complete a series of questions. We ask the same battery of questions used in Step III to gather perceptions of the ads pertaining to different stages of the purchase funnel and brand personality dimensions. But, instead of having participants evaluate existing ads for the purpose of training the generative AI, we now have participants complete the survey as a means of evaluating the generated visuals. This enables us to compare the AI-generated ads with ads that were produced using human labor and the typical creative workflow.

As our proposed workflow only creates visual content and does not produce text, the generated visual content still requires a human touch to complement it with writing and positioning text on top of the ad visual. For this purpose, the best performing ads of step VI can be selected as templates for subsequent refinement.

To illustrate the results of the prompt training, Figure 3 contains a selection of exemplar prompts, c, consisting of AIDA funnel phases, the product, and a specific brand personality dimension, comparing them to actual ads collected from the Internet.

Figure 3: Examples of Actual and Generated Ads with Trained Prompts



Empirical Analyses

We conduct a series of studies in which we demonstrate the capabilities of our proposed generative AI workflow. In Study 1, we evaluate the capability of generative AI to produce visual content that is aligned with measures associated with different stages of the purchase funnel. In Study 2, we examine how these results compare to an alternative strategy of relying on existing prompts of the GDM related to objective image content, rather than our approach of training on mindset metrics. Since ads are usually geared towards a combination of various marketing objectives by brands, study 3 evaluates if our algorithm can be used to generate visuals that successfully evoke specific brand personality dimensions, as well as combinations of multiple dimensions and combinations of brand personality and funnel metrics. Finally, to investigate whether existing brand equity might drive our results, study 4 replicates our analyses with an automotive brand unfamiliar to the target group.

Study 1: Generating Content for Different Stages of the Purchase Funnel

Method

In this first study, we apply the workflow presented in Figure 1 to demonstrate that our approach is capable of generating ads that align with the different stages of the purchase funnel. For each stage of the purchase funnel (AIDA), we generate 10 ads based on the workflow. The set of actual car ad images used for training corresponds to the best performing ads based on the survey question associated with that particular stage of the purchase funnel. Accordingly, we generate 10 ads designed for high performance on each of the following measures: 1) attention, 2) interest, 3) likability (desire), and 4) purchase intention (action). In addition to these 40 generated ads, we also generate 10 ads based on a model trained on the best-rated images according to the mean across all AIDA-based purchase funnel phases (all phases). Each study participant rated 10 of these ads per survey, and all available ads are rated 10 times on average.

We employ Prolific to recruit 571 study participants (279 females, 290 males, 2 preferred not to state their gender, $\text{age}_{\text{range}} = 25 - 75$). All 571 participants have a driving license and are thus potential consumers. Participants were randomly assigned to provide responses to purchase funnel metrics questions on a Likert scale ranging from 1 to 7 regarding attention ("This advertisement would stand out in comparison to other advertisements"), interest ("I find the product in this advertising interesting"), desire ("I like the product in this advertisement"), and purchase action ("If I were in the market for a car right now, I would buy the car in this advertisement") (Smith et al. 2008). Participants who did not pass an attention check were removed from the study (see Appendix B for details). Survey participants were confronted with 10 ads, randomly selected from a set of the 559 ads that includes 10 actual Polestar 3 ads, 499 actual competitive car ads, and 50 ads using the AI workflow. This results in a total of 21,998 ratings for these 559 images.

Results and discussion

Table 1 contains the average ratings for each set of ads. We find that generated ads, trained on each phase of the purchase funnel, perform better at that stage than both the average performance of actual competitors' ads and the actual Polestar ads. Interestingly, this finding holds regardless of the purchase funnel phase on which we train the generative AI. On average, the generated content always exceeds the content we found on the Internet. Note that brands have paid to create and distribute all of these ads. Comparing all generated ads across all phases of the purchase funnel, we do not detect statistically significant differences. In Table 1, we therefore focus the statistical testing on the comparison with generated ads trained on the average of all funnel phases (all phases).

For attention and interest, we find a difference of approximately one scale point on the seven point scale ($p < .01$), i.e., $> 14\%$ higher KPI performance in favor of the generated ads compared to both the average of the actual competitors' ads and the average of the actual Polestar ads. For likability and purchase intent, the difference compared to all competitors' ads is smaller, but still favors the generated ads and still exceeds half a scale point, i.e., $> 7\%$

Table 1: Consumer Ratings of Generated and Actual Ads

Dimension	Group	Descriptives M	Descriptives SD	Welch's t-test t
Attention	Trained on All Phases	4.68	1.66	
	Actual Competitive Ads	3.77	1.75	5.44**
	Actual Polestar Ads	3.43	1.69	5.28**
	Trained on Attention	4.47	1.76	.87
	Trained on Interest	4.71	1.45	-.11
	Trained on Likability	4.54	1.61	.56
	Trained on Purchase Intent	4.47	1.66	.89
Interest	Trained on All Phases	4.81	1.62	
	Actual Competitive Ads	3.88	1.82	5.66**
	Actual Polestar Ads	3.82	1.83	4.05**
	Trained on Attention	4.71	1.68	.44
	Trained on Interest	4.69	1.60	.55
	Trained on Likability	4.65	1.68	.63
	Trained on Purchase Intent	4.43	1.67	1.63
Likability	Trained on All Phases	4.80	1.73	
	Actual Competitive Ads	4.14	1.79	3.79**
	Actual Polestar Ads	4.27	1.60	2.25*
	Trained on Attention	4.78	1.55	.07
	Trained on Interest	4.82	1.71	-.10
	Trained on Likability	4.88	1.83	-.31
	Trained on Purchase Intent	4.75	1.65	.21
Purchase Intent	Trained on All Phases	4.15	2.07	
	Actual Competitive Ads	3.40	1.81	3.57**
	Actual Polestar Ads	3.66	1.93	1.78†
	Trained on Attention	4.05	1.93	.37
	Trained on Interest	4.10	1.94	.18
	Trained on Likability	4.40	1.93	-.86
	Trained on Purchase Intent	4.03	2.03	.40

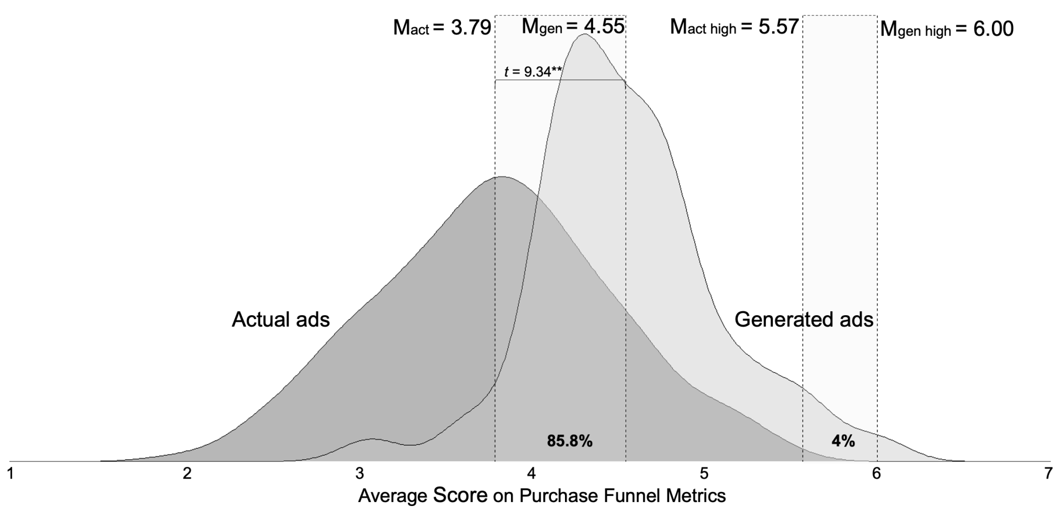
†$p \leq .10$; *$p \leq .05$; **$p \leq .01$.

higher performance on mindset metrics ($p < .01$). With regard to the actual Polestar ads, we find a similar effect size. However, on purchase intent, the difference is only marginally significant ($p < .10$).

The actual ads produced by Polestar and its competitors that we collected from the Internet involved human creativity, photo shooting and image processing. The generated ads, on the other hand, involved no further editing or additional prompting and were directly taken from the output of the generative AI workflow. However, the values presented in Table 1 are only the means of underlying distributions. How do the best performing actual ads on the Internet compare to ones from generative AI? Can one pick generated ads at random, or would that risk selecting ads that fall below the average of what is currently used? To investigate these questions, we compare the distribution of all actual ads with the distribution of all generated ads (see Figure 4).

This results in several noteworthy conclusions. First, comparing the generated ads to the actual competitors' ads, the average of the generated ads turns out to be better than 85.8% of all actual competitors' ads, i.e., selecting a generated ad at random would have an $> 85\%$ probability of exceeding the average performance of human-generated ads in our data. When we turn to the best performing ads, we find that the best performing actual ad still falls below the best performing generated ad ($M_{\text{highest actual}} = 5.57$ vs. $M_{\text{highest generated}} = 6.00$). More specifically, 4% of the generated ads perform better than the best ad that was actually paid for and distributed. We arrive at similar conclusions regarding the distribution of actual Polestar ads (see Appendix C).

Figure 4: Distribution of Average Consumer Ratings of Actual vs. Generated Ads



Since we collected ads available on the Internet, some of these have been distributed earlier than others. The GDM may have picked up more recent advertising trends, so these AI-generated ads may have appeared more contemporary to respondents. Note, however, that more than 90% of the actual ads started distribution after 2020. For this short timeframe, the correlation between time of appearance and average ad ratings is low ($\tau_b < .10$), making this an unlikely driver of results. Moreover, if results were driven by recency alone, then we would expect the best performing (more recent) actual ads to have comparable or superior performance to the best performing generated ad.

Overall, this study suggests that there is potential in generating ads by training visual generative AI on marketing objectives. However, an alternative strategy would be to avoid all fine-tuning and model training and to simply rely on what is directly available in off-the-shelf generative AI, i.e., to use existing prompt language. We investigate the potential of such an approach next.

Study 2: Training New Prompts based on Mindset Metrics vs. Engineering Existing Prompts

There are many creative ways to prompt off-the-shelf generative AI to produce new visual content for advertisements. Arguably, the most popular approach is prompt engineering, where users iterate different prompts until the generative AI yields an output they consider adequate. The success of such an approach depends, among other things, on the human creativity and prompting. For a fair comparison with our human-free approach, we examine a human-free prompt engineering. To find adequate prompts without human intervention, we make use of an image-to-text model, CLIP Interrogation (pharmapsychotic 2022). The CLIP Interrogator transforms images into the single textual description that best represents the image. Put differently, it can be thought of as the reverse of a text-to-image generative AI. We use the resulting description of an image as a text prompt in our GDM to create new web advertisements. While the CLIP Interrogator generates a single textual description, the model produces images probabilistically and therefore can create many alternative images based on the same text prompt. In this study, we select the first image generated by the GDM to test the performance of the CLIP approach against our workflow.

Specifically, we use the ads that performed best across all purchase funnel stages as the input for the CLIP Interrogator. These best performing ads are the same ads that we used to train the generative AI in Study 1. The CLIP Interrogator transformed these images into textual descriptions, one for each ad, which served as a prompt that was fed into the GDM to generate new ads (see Figure 8 in Appendix D).

Method

We recruited 101 participants (49 females, 52 males, ranging 25–75 years in age, all having a driving license) through Prolific. Each respondent rated 10 ads, and each ad was rated 10 times on average in a similar survey as in study 1. We compare 30 ads generated using the prompts produced by CLIP, 30 ads generated to perform well on all phases (as in Study 1) and 40 randomly selected actual ads obtained from the Internet. This results in a total of 4,230 ratings on these 100 images. Ratings were again collected on standard seven-point Likert scales for each of the four funnel phases, as in Study 1. Again, none of the generated ads were manually edited after being produced by the generative AI.

Results and discussion

The comparison of the generated ads based on CLIP prompts with the actual ads results in an inconsistent pattern (Table 2). For purchase intent, the CLIP prompts perform slightly worse than the average of the actual ads, although this difference is not statistically significant. For all other phases, the CLIP approach results in more favorable average ratings than the actual ads ($p < .01$). Apparently, even a relatively simple, low effort approach to prompting can produce novel versions of previous image content that performs comparably or even better than the original ads.

More importantly, when comparing the CLIP approach to our AI workflow of training new prompts based on consumer responses, we find CLIP performs worse at all stages of the purchase funnel, with differences ranging between .4 and a one full scale-point ($p < .01$). Our workflow again outperforms actual ads, reproducing the results from Study 1. This suggests

Table 2: Performance of Ads Generated with Existing and New Prompts vs. Actual Ads

Dimension	Group	Descriptives		Welch's t-test
		M	SD	t
Attention	Trained on All Phases	4.73	1.52	
	Actual Competitive Ads	3.53	1.66	10.19**
	Generated with CLIP	4.30	1.60	3.52**
Interest	Trained on All Phases	4.83	1.61	
	Actual Competitive Ads	3.57	1.83	9.90**
	Generated with CLIP	4.21	1.78	4.63**
Likability	Trained on All Phases	4.91	1.57	
	Actual Competitive Ads	3.87	1.77	8.36**
	Generated with CLIP	4.15	1.79	5.62**
Purchase Intent	Trained on All Phases	4.54	1.81	
	Actual Competitive Ads	3.35	1.91	8.56**
	Generated with CLIP	3.33	1.96	8.05**

**$p \leq .01$.

training on marketing objectives is useful and promising, and it is beneficial to move beyond off-the-shelf applications of general purpose generative AI.

Taken together, Studies 1 and 2 demonstrate the potential for a generative AI that has been trained with consumer response data on specific marketing outcomes to outperform actual advertisements that have been distributed (and paid for) in terms of funnel metrics. However, does this come at the expense of how the brand is perceived by consumers after seeing the ad? Perhaps the ads generated using AI attract more attention to the brand, but these ads may not be capable of conveying the desired brand perception in the minds of consumers. We therefore investigate if it is possible to convey brand personality traits (in our empirical context, "ruggedness" and "luxury") using generative AI *and* if this can be done while performing well on purchase funnel metrics.

Study 3: Incorporating Brand Personality

In this study, we examine the potential for generative AI to produce visual content for ads that are capable of evoking specific perceptions of the brand. Our approach borrows from research by Liu et al. (2020), who investigate brand perceptions based on both user- and firm-generated content.

Method

267 study participants (118 females, 149 males, $Age_{range} = 25–75$, all have a driving license) acquired via Prolific were assigned to an online survey in which they were asked to rate 10 randomly chosen ads on seven-point Likert scales to determine whether they perceived the ads as conveying "ruggedness" ("This advertisement looks rugged to me") and "luxury" ("This advertisement looks luxurious to me") as well as the purchase funnel metrics from the previous studies and the same attention check. Each ad was rated 10 times on average.

In contrast to Studies 1 and 2, we train the GDM on brand personality traits using images selected from Flickr. Following Liu et al. (2020), we use the Flickr images to infuse our GDM with the desired brand personality trait. In so doing, we combine the visual language of automotive display ads with the desired brand personality traits. In total, we randomly sampled 155 actual competitive ads, 10 actual Polestar ads, 10 generated ads trained on all funnel phases, 10 ads generated to portray "ruggedness", 10 trained on "luxury", and 10 trained in addition on both "ruggedness" and "luxury". This results in a total of 205 images and 12,822 individual ratings (see Figure 3 for generated images geared towards ruggedness and luxury).

Results and discussion

We first study whether it is possible to steer a GDM to represent brand personality associations in images. Table 3 summarizes the results. For the brand personality dimension "ruggedness", we find that training a GDM on ruggedness indeed results in images rated highest on that dimension. Specifically, respondents rated these images as 4.9 on a seven-point scale, which is significantly different from the scale midpoint ($p < .05$), higher than the average of all actual competitive ads ($M_{actual} = 3.71$, $p < .01$), higher than the actual

Polestar ads ($M_{polestar}$ = 4.07, p < .01), and more than one scale point higher than a model trained on no brand dimension ($M_{no\ dimension}$ = 3.53). Interestingly, a model trained exclusively on another brand personality dimension, "luxury", results in the lowest "ruggedness" ratings we observe (p < .01).

In terms of "luxury" perceptions, we find it is also possible to train that personality dimension. Specifically, generated ads geared towards luxury are rated 5.43, which is also significantly above the scale midpoint as well as all other alternative actual and generated ads. In terms of effect size, actual competitive ads (M_{actual} = 4.01) and actual Polestar ads ($M_{polestar}$ = 4.5) convey about a scale point less luxury to respondents (both p < .01). Furthermore, no training on brand personality leads to lower luxury perceptions ($M_{no\ dimension}$ = 5.00, p < .05). A model trained on "ruggedness" performs the worst on "luxury" (M_{rugged} = 3.72, p < .01).

Table 3: Driving Brand Image with Generative AI

Dimension	Group	Descriptives		Welch's t-test	
		M	SD	*t*	
				Trained on Rugged	Trained on Rugged & Luxury
Rugged	Trained on Rugged	4.90	1.68		-2.15*
	Trained on Rugged & Luxury	4.39	1.75	2.15*	
	Actual Competitive Ads	3.71	1.86	6.96**	3.87**
	Actual Polestar Ads	4.07	1.66	3.55**	1.35
	Trained without Brand Dimension	3.53	2.08	5.26**	3.26**
	Trained on Luxury	3.11	1.81	7.34**	5.21**
				Trained on Luxury	Trained on Rugged & Luxury
Luxury	Trained on Luxury	5.43	1.39		-4.84**
	Trained on Rugged & Luxury	4.46	1.52	4.84**	
	Actual Competitive Ads	4.01	1.80	9.90**	2.88**
	Actual Polestar Ads	4.50	1.51	4.59**	-.20
	Trained without Brand Dimension	5.00	1.43	2.22*	-2.67**
	Trained on Rugged	3.72	1.78	7.72**	3.22**

*$p \leq .05$; **$p \leq .01$.

This study demonstrates that generative AI can replicate the visual language needed to communicate specific brand personality dimensions. The results also suggest that the concepts of "ruggedness" and "luxury" might be semantically opposing elements, at least in terms of the visual interpretation of our GDM. This might make it particularly challenging to combine both concepts into a single advertisement. According to Table 3, this is indeed the case. Prompting both trained dimensions simultaneously cannot reach the image association levels of a model trained individually on each dimension. Specifically, a model aligned with more complex brand image objectives performs worse on "ruggedness" than a model exclusively aligned with "ruggedness" ($M_{\text{rugged \& luxury}} = 4.39$ vs. $M_{\text{rugged}} = 4.9$, $p < .05$). Similarly, a model trained on both dimensions performs worse on "luxury" than a "luxury"-exclusive model ($M_{\text{rugged \& luxury}} = 4.46$ vs. $M_{\text{luxury}} = 5.43$, $p < .05$). While the simultaneous model does not exceed the image association of the actual Polestar ads ($p > .10$), it does exceed both the average competitive ads ($p < .01$) and the scale midpoint ($p < .05$) for both brand image dimensions. Apparently, more complex brand image associations are attainable even when these are semantically poorly aligned. However, per image dimension, this comes at the expense of what is attainable with an exclusive focus on a single dimension. Note that the Polestar brand is already strongly positioned on the luxury dimension. Further, exceeding that level when combined with a partly contradictory association such as ruggedness appears challenging.

While these results demonstrate the capacity to use generative AI to create content that evokes specific brand personality dimensions, does doing so hamper performance on purchase funnel objectives? Table 4 summarizes the results on these metrics. When comparing the different generated ads, we find no significant differences. As in Studies 1 and 2, each set of AI-generated images performs better across all purchase funnel metrics than the average ratings of all competitive ads ($p < .05$). Compared to the actual Polestar ads, however, the results are mixed. While we find that the AI-generated ads outperform the Polestar ads early in the purchase funnel (attention and interest) ($p < .05$), mixing brand personality with funnel objectives yields ads that perform statistically equivalently for likability and purchase intent.

We conclude from this study that it is also possible to train generative AI to convey a chosen brand personality dimension without compromising the performance on purchase

Table 4: Funnel Performance of Generated Ads When Prompting both Funnel and Brand Image Objectives for Polestar

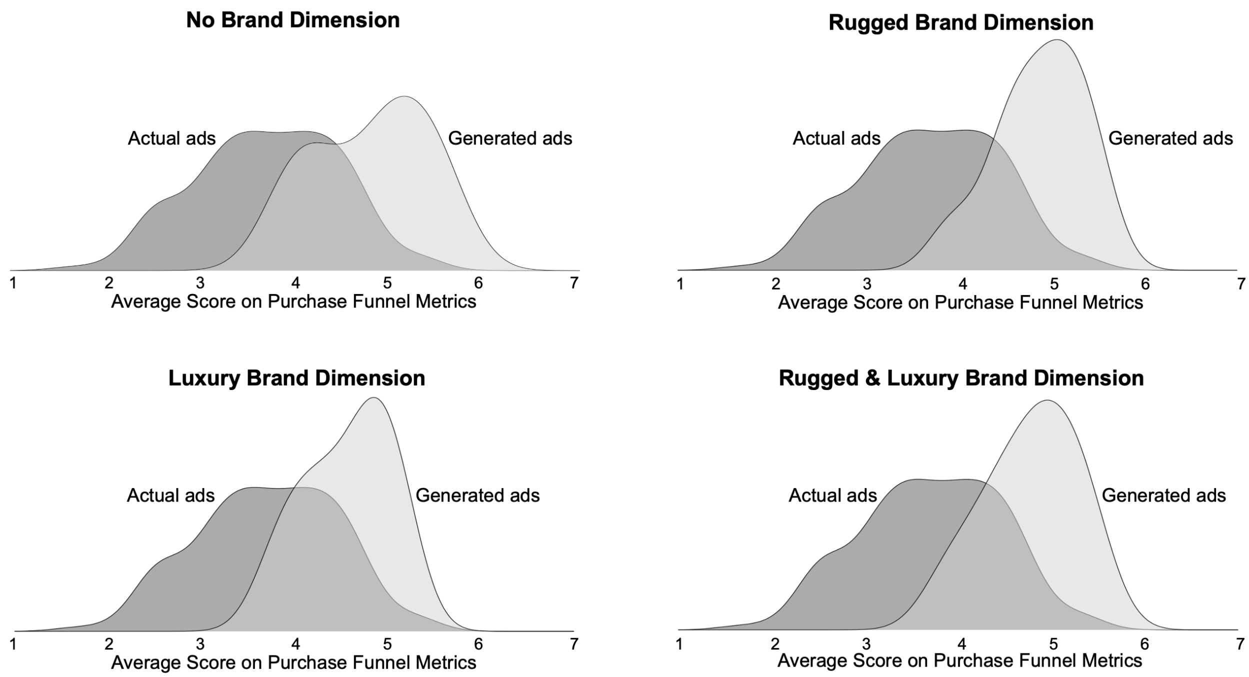
Dimension	Group	Descriptives		Welch's t-test	
				t	
		M	SD	Actual Competitive Ads	Actual Polestar Ads
Attention	Actual Competitive Ads	3.46	1.79		2.38*
	Actual Polestar Ads	3.86	1.64	-2.38*	
	Trained without Brand Dimension	4.57	1.73	-6.37**	-3.00**
	Trained on Rugged	4.99	1.47	-10.16**	-5.17**
	Trained on Luxury	4.43	1.81	-5.34**	-2.37**
	Trained on Rugged & Luxury	4.53	1.61	-6.59**	-2.96**
Interest	Actual Competitive Ads	3.82	1.86		2.34*
	Actual Polestar Ads	4.25	1.80	-2.34*	
	Trained without Brand Dimension	5.04	1.75	-6.93**	-3.18**
	Trained on Rugged	5.13	1.56	-8.18**	-3.70**
	Trained on Luxury	4.84	1.67	-6.01**	-2.41*
	Trained on Rugged & Luxury	4.92	1.71	-6.39**	-2.74**
Likability	Actual Competitive Ads	4.11	1.79		3.23**
	Actual Polestar Ads	4.69	1.66	-3.32**	
	Trained without Brand Dimension	5.19	1.75	-6.16**	-2.08*
	Trained on Rugged	5.05	1.50	-6.09**	-1.60
	Trained on Luxury	4.79	1.66	-4.02**	-.42
	Trained on Rugged & Luxury	4.99	1.57	-5.53**	-1.31
Purchase Intent	Actual Competitive Ads	3.45	1.98		5.39**
	Actual Polestar Ads	4.43	1.75	-5.39**	
	Trained without Brand Dimension	4.45	2.02	-4.94**	-.08
	Trained on Rugged	4.34	1.94	-4.49**	.35
	Trained on Luxury	4.17	2.02	-3.53**	.97
	Trained on Rugged & Luxury	4.60	1.95	-5.84**	-.66

*$p \leq .05$; **$p \leq .01$.

funnel metrics. As such, the long-term strategic brand positioning need not come at the expense of more short-term transactional goals. It is even possible to mix ostensibly conflicting brand associations.

As in study 1, we examine the distribution of the average ratings across funnel phases of the generated ads including the brand dimensions and compare these to the distributions of the actual competitive ads from this study (see Figure 5). It can be seen that the generated ads, irrespective of the added brand dimension, show similar results relative to actual ads as in Study 1. Specifically, all generated ads including no brand dimension, rugged or luxury brand dimension, or both brand dimensions all show similar performance on the funnel

Figure 5: Distribution of Average Consumer Ratings of Actual vs. Generated Ads for Individual Brand Dimensions



phases, and all outperform actual competitors' ads ($M_{\text{actual}} = 3.71$, $M_{\text{no dimension}} = 4.80$, $M_{\text{rugged}} = 4.88$, $M_{\text{luxury}} = 4.47$, $M_{\text{rugged \& luxury}} = 4.78$).

Thus far, we focus our analysis on a single brand and product. Specifically, we chose Polestar as a relatively young but also already popular brand. Across Studies 1-3, respondents perceived the Polestar advertisements as more engaging than the market average. It is conceivable that the favorable position of the Polestar brand may have contributed to more favorable assessments of novel Polestar advertisements. To assess the robustness of our approach and rule out this potential explanation, we next study a brand that is unknown in the region from which our respondents are drawn and for which they do not have any predetermined attitudes or consumer associations.

Study 4: Leveraging Visual Generative AI for an Unknown Brand

To leverage the previous training of the automotive market and to make our results comparable with the previous studies, we select another electric vehicle brand with which respondents are not familiar. Specifically, we study the Chinese car brand, Nio, that is not yet available in the market but is interested in entering.

Method

We use the previously described procedure and ask 229 participants with a driving license (117 females, 111 males, 1 preferred not to state the gender, $\text{age}_{\text{range}} = 25\text{–}75$), acquired over Prolific, to rate 170 actual ads, 40 generated ads, and 5 actual Nio ads on purchase funnel metrics and perceptions of brand personality. Each respondent rated 10 ads, and each ad is rated 10 times on average. This results in a total of 13,530 individual ratings on 215 images.

Results and discussion

In terms of brand funnel performance, the results presented in Table 5 exhibit a highly consistent pattern with Study 3. Generated ads perform better than the average of all competitors' ads and better than all Nio ads from the Chinese market ($p < .05$). Training on brand image associations also does not appear to inhibit funnel performance, as we see no systematic pattern in favor of models trained exclusively on funnel metrics. This is encouraging and suggests that the potential of generative AI in terms of mindset metrics is not driven by peculiarities of the Polestar brand or consumers associations with that particular brand.

Table 5: Funnel Performance of Generated Ads When Prompting both Funnel and Brand Image Objectives for an Unknown Brand

Dimension	Group	Descriptives		Welch's t-test t	
		M	SD	Actual Competitive Ads	Actual Nio Ads
Attention	Actual Competitive Ads	3.52	1.86		1.08
	Actual Nio Ads	3.77	1.67	-1.08	
	Trained without Brand Dimension	4.83	1.59	-7.81**	-3.86**
	Trained on Rugged	5.02	1.62	-9.26**	-4.59**
	Trained on Luxury	4.47	1.68	-5.60**	-2.52*
	Trained on Rugged & Luxury	4.67	1.78	-6.42**	-3.19**
Interest	Actual Competitive Ads	3.92	1.88		1.13
	Actual Nio Ads	4.20	1.82	-1.13	
	Trained without Brand Dimension	5.09	1.47	-7.51**	-3.13**
	Trained on Rugged	5.17	1.56	-8.06**	-3.42**
	Trained on Luxury	5.04	1.62	-6.87**	-2.90*
	Trained on Rugged & Luxury	5.05	1.58	-7.09**	-2.96**
Likability	Actual Competitive Ads	4.15	1.86		1.58
	Actual Nio Ads	4.50	1.64	-1.58	
	Trained without Brand Dimension	5.31	1.45	-7.54**	-3.07**
	Trained on Rugged	5.25	1.47	-7.46**	-2.87**
	Trained on Luxury	5.17	1.57	-6.47**	-2.51*
	Trained on Rugged & Luxury	5.07	1.69	-5.43**	-2.07*
Purchase Intent	Actual Competitive Ads	3.37	2.02		.88
	Actual Nio Ads	3.61	1.94	-.88	
	Trained without Brand Dimension	4.49	1.93	-5.50**	-2.71**
	Trained on Rugged	4.55	1.78	-6.65**	-3.04**
	Trained on Luxury	4.24	2.08	-4.20**	-1.93†
	Trained on Rugged & Luxury	4.38	2.06	-4.87**	-2.35*

†$p \leq .10$; *$p \leq .05$; **$p \leq .01$.

With regard to brand image, creating associations for unknown brands appears to be more challenging. We also find indications that "ruggedness" and "luxury" involve opposing elements. Training both simultaneously reduces performance on each personality dimension, in particular on "ruggedness". Ads geared towards "ruggedness" and "luxury" one at a time perform systematically better on the respective personality association than both the average of the market and the actual Nio ads (see Table 6). However, on both dimensions, ads only perform directionally better than GDM ads lacking additional brand image training. The distribution of these ad evaluations results in similar conclusions (See Appendix E). The findings of this study suggest GDMs are useful for building a novel brand image, but this process may require repeated exposure and longer-term effort than reinforcing existing associations of more established brands.

Table 6: Brand Image Performance of Visual Gernative AI for an Unknown Brand

Dimension	**Group**	**Descriptives**		**Welch's t-test**	
		M	SD	*t*	
				Trained on Rugged	Trained on Rugged & Luxury
Rugged	Trained on Rugged	4.25	1.73		-2.35**
	Trained on Rugged & Luxury	3.70	1.70	2.35*	
	Actual Competitive Ads	3.55	1.83	4.05**	.84
	Actual Nio Ads	3.02	1.29	5.15**	2.86**
	Trained without Brand Dimension	3.82	1.62	1.81†	-.53
	Trained on Luxury	3.30	1.74	4.02**	1.69†
				Trained on Luxury	Trained on Rugged & Luxury
Luxury	Trained on Luxury	5.19	1.47		-.87
	Trained on Rugged & Luxury	5.01	1.51	.87	
	Actual Competitive Ads	4.11	1.43	7.26**	5.91**
	Actual Nio Ads	4.57	1.41	2.60*	1.83†
	Trained without Brand Dimension	5.03	1.43	.76	-.11
	Trained on Rugged	4.50	1.53	3.39**	2.48*

†$p \leq .10$; *$p \leq .05$; **$p \leq .01$.

Discussion

Visual online advertising is a mainstay in today's digital marketing landscape for brand building and driving sales. However, the process of crafting visual ad content is complex and time-consuming, often requiring many iterations among creatives, visual designers and marketing managers. To understand the potential of alternative advertising concepts, marketers typically survey consumers with potential ad concepts on various KPIs, as well as the desired mindset associations. This can result in costly iterations between marketers, agencies, and market researchers to find creative ways to achieve the desired objectives.

While brands have expressed interest in the use of generative AI technologies, we know little about the suitability of these technologies in support of a specific marketing outcome. Conversations with agencies suggest that interest in the generative capabilities of AI for marketing are driven primarily by the potential cost savings (e.g., prompting a suitable location for a photo shoot instead of finding one, travelling there and taking photos). But, are cost savings the only rationale for exploring the use of generative AI in creating marketing content? Or can generative AI be used to create content that performs well in the eyes of consumers?

In this paper, we have presented and tested a novel workflow for the creation of visual content for advertisements that is "fine-tuned" for specific marketing objectives. We guide the performance of an open-source image generative AI with the objectives of both performing well on common measures of ad performance throughout the purchase funnel and evoking specific brand personality dimensions. Using two different automotive brands that vary in their level of familiarity among consumers, we demonstrate generative AI models are capable of outperforming actual industry ads as well as existing branded ads in terms of key performance metrics related to the purchase funnel. Moreover, we find that aligning with additional brand associations such as "luxury" and "ruggedness" does not inhibit funnel performance. In addition, we illustrate that current AI models are capable of combining multiple objectives, generating ads that are highly effective in terms of short-term performance metrics while still conveying a desired brand's personality.

Our insights are of value to both researchers and practitioners. On the theoretical side, we demonstrate the suitability of AI-powered systems to generate ads that are optimized for abstract properties like "attention" or "likability." In this way, visual generative AI found its own creative ways to attain marketing outcomes. We deliberately did not prompt our GDM what to put on images (no prompts such as houses, rocks, streets, etc.) or engage in other forms of prompt engineering. Rather, our approach relies entirely on the capabilities of finding commonalities between effective ads that are useful in producing entirely new ads.

Interestingly, when we compare the best performing generated ads to the best performing actual ads that we used for model training, we find that the generated ads outperform their training material significantly in terms of the average score across all funnel phases ($M_{training} = 4.72$, $M_{generated} = 5.38$, $t = 7.53$, $p < .01$). For a more detailed table, see Table 7 in Appendix F. In terms of mindset metrics, GDMs can apparently accomplish more than mere replication and produce more effective advertising content than the data it is trained on.

How is this possible? One potential reason is that many GDMs are trained on image data sets such as LAION that contain image aesthetics. This allows GDMs to produce aesthetically pleasing images. Using this as a foundation and combining it with input on marketing KPIs yields high-quality visual advertising content that can even surpass its advertising training material.

Another reason could be rooted in fluency theory. The performance of generative AI might owe to the fact that these models are good at averaging, and averages have been shown to be perceptually pleasing. We tested if this hypothesis holds by executing a study where we compared the fluency of all generated ads to all actual ads (see Appendix G). Indeed, we find that the generated ads are perceived as more fluent than the average of all actual ads in our database and also the ones we trained on ($p < .01$). This suggests training generative AI on marketing objectives can produce more fluent content, which may explain the good performance of visual generative AI. While this is not the primary focus of the current research, we believe it is promising to conduct further research on the perceptions of generated content compared to the source material.

Given the apparent quality of its output, we note that the proposed workflow can be used by researchers to create stimuli for their research. Our use of open-source software makes the workflow accessible to researchers who may have a need to create visual content that evokes particular brand perceptions among consumers.

Another direction that warrants future research is testing the proposed workflow in other empirical settings. While we have tested the performance of the model for two brands in a single industry to demonstrate its potential, future work may consider its performance in different industries, particularly in industries in which purchase decisions are less involved such as FMCGs. The current research is also limited in that we gauge consumers' response via popular survey measures. As brands start distributing AI-generated images at a larger scale, it may be possible to track long-term consequences on brand equity or even overall firm performance. Future work could explore the capability of AI-generated visuals to these outcomes.

From a practical standpoint, our findings suggest that the creative process of marketing that involves entire industries (creative advertising agencies, media agencies, advertising market research, marketing consultants) may undergo fundamental shifts in the coming years. Marketers can easily produce an abundance of ad content that can be geared to any conceivable KPI or brand association. Finding creative solutions to attain these KPIs is no longer a bottleneck, allowing for varied concepts to be tested with consumers. This can potentially support increased targeting and personalization, as content creation costs and the time associated with it can be significantly reduced. Agile marketers are also interested in quickly responding to current developments. GDMs can circumvent traditional time-consuming image production, making them a useful tool when quick responses are essential. Once GDMs are trained on the desired marketing objectives, additional prompting on objective image content can add links to current events (e.g., sports, political, entertainment events). As we demonstrate through our workflow, a critical element is the incorporation of the marketing objective into the GDM. Rather than striving to build larger GDMs, we encourage researchers to consider how their objectives could be incorporated into existing models.

Could the widespread adoption of generative AI make advertising homogeneous and less differentiated? The results on fluency point to that direction. Whether this will happen will depend on how marketers employ generative AI. One way for brands to differentiate themselves from competitors in such a media landscape is to inject desired brand personality dimensions, as we have demonstrated. Yet, brands may seek similar positioning (e.g., luxury for automotive brands). This may make it necessary to utilize additional creative efforts such as Coca-Cola is seeking or additional photo editing. Alternatively, brands can produce effective ads in traditional ways and leverage proprietary training material for visual generative AI. Brands with excellent communication skills might be in the position to leverage these to produce many new versions that attain similar objectives for different channels and target groups. Conversely, brands with smaller resources or less success in connecting with consumers can tap into the communication quality of stronger competitors. This research has shown that both developments are technically possible. It will depend on corporate adoption and available (software) services, whether stronger or trailing brands stand more to gain.

A related question of importance will be the nature of client and agency relationships going forward. While clients may embrace generative AI "in-house", which may pose something of a threat to agencies, agencies are also examining the capabilities of these tools. Agencies may find themselves able to take on more clients without expanding their ranks thanks to the time-savings afforded by generative AI. Agencies could also use generative AI as an input to further optimize, or could add creative prompts as in our introductory examples to further improve and differentiate their generated ads. We deliberately did not edit or use any further human creativity in our studies. However, one area that warrants additional research is exploring the potential interplay of generative AI and human creativity to achieve both short-term business objectives and distinctiveness in the minds of consumers.

In addition to agencies, new players that have not traditionally been producing advertisements are entering the market. This includes major online advertising channels such as Google and Meta having announced services in that direction, but also traditional media channels that could conceivable leverage the advertising content available to them to train generative AI in similar ways as we have done. This would make it possible to move beyond ad distribution and enter ad production. While it is not clear who will be the winner in this race, it is clear that brands will have more options to produce marketing content than was conceivable only a few years ago.

Our approach combines three different sets of images: images of the product to be featured in the ad, actual ads from which the general visual language of the category is learned, and Flickr images to characterize the visual language associated with dimensions of brand personality. For agencies and advertising channels that have access to large sets of actual ads, brand personality dimensions could potentially be trained using actual ads rather than Flickr images. For our data, this approach was less effective than using non-advertising training data. However, regarding brand personality, in particular, cross-industry learning might be possible. For example, closely related industries (e.g., motorcycles and cars) or industries with similar positioning (e.g., luxury watches and cars) may provide additional valuable training data to strengthen brand image associations. It would be interesting to explore the benefits of larger databases of advertising assets than available in this research.

Another opportunity for future work is to examine the source of consumer feedback. Our AI workflow incorporates consumer input in the form of survey responses to prior ads. This critical step enables us to identify ads that are successful for a particular business outcome, and then infuse the generative AI with the visual language of these successful ads. One could envision an active learning framework in which advertisers experiment in near real-time with AI-generated ads, deploying them online and gathering feedback in the form of click-through rates to cull the wheat from the chaff. We leave the examination of other sources of consumer feedback to power generative AI to future research.

While we are optimistic about the opportunities afforded by generative AI, it is not without its limitations. Just as research continues to explore the potential applications within marketing, it will be critical to identify those areas to which it is not (yet) well suited. It may very well be the case that generative AI is akin to an eager intern, capable of capturing the broad strokes of content creation, but lacking the finesse to tailor the output to a particular brand or application. Moreover, there are important questions to be answered as far as how receptive consumers will be to content created using AI systems, and the likely need for regulations governing aspects of what can be used for training and what must be disclosed to consumers when such systems are used. We urge the marketing community to take an active role in such discussions.

Our analyses demonstrate that generative AI can be a useful tool to efficiently support strategic brand objectives. But, marketers must still decide how they want their brands to be seen in the minds of consumers. This will come down to finding the right strategy and developing the best possible vision for a brand. Once that has been determined, generative AI may prove to be a useful tool for producing content that conveys such meaning to consumers.

We hope our work encourages further research in these and related directions.

References

Aaker, Jennifer L. (1997), 'Dimensions of brand personality', *Journal of marketing research* **34**(3), 347–356.

Affonso, Felipe M. and Chris Janiszewski (2023), 'Marketing by design: The influence of perceptual structure on brand performance', *Journal of Marketing* p. 00222429221142281.

Batra, Rajeev and Kevin Lane Keller (2016), 'Integrating marketing communications: New findings, new lessons, and new ideas', *Journal of marketing* **80**(6), 122–145.

Brand, James, Ayelet Israeli and Donald Ngwe (2023), 'Using gpt for market research', *Available at SSRN 4395751* .

Brasel, S. Adam and James Gips (2008), 'Breaking through fast-forwarding: Brand information and visual attention', *Journal of Marketing* **72**(6), 31–48.

Ceylan, Gizem, Kristin Diehl and Davide Proserpio (2023), 'Words meet photos: When and why photos increase review helpfulness', *Journal of Marketing Research* **0**(0).

Chaudhuri, Arjun and Morris B. Holbrook (2001), 'The chain of effects from brand trust and brand affect to brand performance: the role of brand loyalty', *Journal of marketing* **65**(2), 81–93.

Colicev, Anatoli, Ashwin Malshe, Koen Pauwels and Peter O'Connor (2018), 'Improving consumer mindset metrics and shareholder value through social media: The different roles of owned and earned media', *Journal of Marketing* **82**(1), 37–56.

Costello, John P., Jesse Walker and Rebecca Walker Reczek (2023), '"choozing" the best spelling: Consumer response to unconventionally spelled brand names', *Journal of Marketing* p. 00222429231162367.

Cramer-Flood, Ethan (2023), 'Worldwide digital ad spending 2023', *eMarketer* .
URL: *https://content-na1.emarketer.com/worldwide-digital-ad-spending-2023*

Dall'Olio, Filippo and Demetrios Vakratsas (2023), 'The impact of advertising creative strategy on advertising elasticity', *Journal of Marketing* **87**(1), 26–44.

Datta, Hannes, Kusum L. Ailawadi and Harald J. van Heerde (2017), 'How well does consumer-based brand equity align with sales-based brand equity and marketing-mix response?', *Journal of Marketing* **81**(3), 1–20.

Dew, Ryan, Asim Ansari and Olivier Toubia (2022), 'Letting logos speak: Leveraging multiview representation learning for data-driven branding and logo design', *Marketing Science* **41**(2), 401–425.

Diab, Mohamad, Julian Herrera, Musical Sleep, Bob Chernow and Coco Mao (2022), 'Openart: Stable diffuision prompt book'.
URL: *https://openart.ai/promptbook*

Dzyabura, Daria and Renana Peres (2021), 'Visual elicitation of brand perception', *Journal of Marketing* **85**(4), 44–66.

Graf, Laura K. M., Stefan Mayer and Jan R. Landwehr (2019), 'Measuring processing fluency: One versus five items', *Journal of Consumer Psychology* **28**(3), 393–411.

Hartmann, Jochen, Mark Heitmann, Christina Schamp and Oded Netzer (2021), 'The power of brand selfies', *Journal of Marketing Research* **58**(6), 1159–1177.

Heitmann, Mark, Jan R. Landwehr, Thomas F. Schreiner and Harald J. van Heerde (2020), 'Leveraging brand equity for effective visual product design', *Journal of Marketing Research* **57**(2), 257–277.

John, Deborah Roedder, Barbara Loken, Kyeongheui Kim and Alokparna B. Monga (2006), 'Brand concept maps: A methodology for identifying brand association networks', *Journal of Marketing Research* **43**(4), 549–563.

Keller, Kevin Lane (1993), 'Conceptualizing, measuring, and managing customer-based brand equity', *Journal of Marketing* **57**(1), 1–22.

Keller, Kevin Lane and Donald R. Lehmann (2006), 'Brands and branding: Research findings and future priorities', *Marketing science* **25**(6), 740–759.

Kulczynski, Alicia and Margurite Hook (2023), 'Express: Typography talks: Influencing vintage anemoia and product safety perceptions with vintage typography', *Journal of Marketing* **0**(0), 1–51.

Langlois, Judith H. and Lori A. Roggman (1990), 'Attractive faces are only average', *Psychological science* **1**(2), 115–121.

Li, Yiyi and Ying Xie (2020), 'Is a picture worth a thousand words? an empirical study of image content and social media engagement', *Journal of Marketing Research* **57**(1), 1–19.

Liu, Liu, Daria Dzyabura and Natalie Mizik (2020), 'Visual listening in: Extracting brand image portrayed on social media', *Marketing Science* **39**(4), 669–686.

Lovett, Mitchell J., Renana Peres and Ron Shachar (2013), 'On brands and word of mouth', *Journal of marketing research* **50**(4), 427–444.

Malär, Lucia, Harley Krohmer, Wayne D. Hoyer and Bettina Nyffenegger (2011), 'Emotional brand attachment and brand personality: The relative importance of the actual and the ideal self', *Journal of marketing* **75**(4), 35–52.

Oppenlaender, Jonas (2022), The creativity of text-to-image generation, *in* 'Proceedings of the 25th International Academic Mindtrek Conference', pp. 192–202.

Pamuksuz, Utku, Joseph T. Yun and Ashlee Humphreys (2021), 'A brand-new look at you: predicting brand personality in social media networks with machine learning', *Journal of Interactive Marketing* **56**(1), 1–15.

Park, C. Whan, Deborah J. MacInnis, Joseph Priester, Andreas B. Eisingerich and Dawn Iacobucci (2010), 'Brand attachment and brand attitude strength: Conceptual and empirical differentiation of two critical brand equity drivers', *Journal of marketing* **74**(6), 1–17.

Peng, Ling, Geng Cui, Yuho Chung and Wanyi Zheng (2020), 'The faces of success: Beauty and ugliness premiums in e-commerce platforms', *Journal of Marketing* **84**(4), 67–85.

Peukert, Christian, Ananya Sen and Jörg Claussen (2023), 'The editor and the algorithm: recommendation technology in online news', *Management Science* **0**(0), 1–16.

pharmapsychotic (2022), 'Clip interrogator'.
URL: *https://huggingface.co/spaces/pharmapsychotic/CLIP-Interrogator*

Pieters, Rik and Michel Wedel (2004), 'Attention capture and transfer in advertising: Brand, pictorial, and text-size effects', *Journal of marketing* **68**(2), 36–50.

Pieters, Rik, Michel Wedel and Rajeev Batra (2010), 'The stopping power of advertising: Measures and effects of visual complexity', *Journal of Marketing* **74**(5), 48–60.

Pogacar, Ruth, Justin Angle, Tina M. Lowrey, L.J. Shrum and Frank R. Kardes (2021), 'Is nestlé a lady? the feminine brand name advantage', *Journal of Marketing* **85**(6), 101–117.

Proserpio, Davide, John R. Hauser, Xiao Liu, Tomomichi Amano, Alex Burnap, Tong Guo, Dokyun Lee, Randall Lewis, Kanishka Misra and Eric Schwarz (2020), 'Soul and machine learning', *Marketing Letters* **31**(0), 393–404.

Reber, Rolf, Norbert Schwarz and Piotr Winkielman (2004), 'Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience?', *Personality and social psychology review* **8**(4), 364–382.

Reisenbichler, Martin, Thomas Reutterer, David A. Schweidel and Daniel Dan (2022), 'Frontiers: Supporting content marketing with natural language generation', *Marketing Science* **41**(3), 441–452.

Rombach, Robin, Andreas Blattmann, Dominik Lorenz, Patrick Esser and Björn Ommer (2022), High-resolution image synthesis with latent diffusion models, *in* 'Proceedings of the IEEE/CVF conference on computer vision and pattern recognition', pp. 10684–10695.

Ruiz, Nataniel, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein and Kfir Aberman (2023), Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation, *in* 'Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition', pp. 22500–22510.

Smith, Robert E., Jiemiao Chen and Xiaojing Yang (2008), 'The impact of advertising creativity on the hierarchy of effects', *Journal of advertising* **37**(4), 47–62.

Toubia, Olivier and Oded Netzer (2017), 'Idea generation, creativity, and prototypicality', *Marketing science* **36**(1), 1–20.

Troncoso, Isamar and Lan Luo (2022), 'Look the part? the role of profile pictures in online labor markets', *Marketing Science* **0**(0).

Wedel, Michel and Rik Pieters (2000), 'Eye fixations on advertisements and memory for brands: A model and findings', *Marketing science* **19**(4), 297–312.

Zaltman, Gerald and Robin A. Coulter (1995), 'Seeing the voice of the customer: Metaphor-based advertising research', *Journal of Advertising Research* **35**(4), 35–51.

Appendix

Appendix A: Image scraping process

Scraping car ads

Before we started scraping ads, we investigated where these ads are stored. This was done by browsing popular websites based on the top websites of Similarweb (e.g., Yahoo, NY Times, Fox News) and manually inspecting the elements of the website to identify links corresponding to the URL of ads. We found that most Google ads are stored in three leading URLs (`https://tpc.googlesyndication.com/simgad/`, `https://tpc.googlesyndication.com/daca_images/simgad/`, `https://s0.2mdn.net/simgad/`), followed by a numeric combination between 15 and 20 numbers. Using the web archive (`http://web.archive.org/cdx/search/cdx?url=`), we were able to trace all possible links where ads were stored. In total, we found 2.5 million ad links. After filtering for only unique ads from April 2012 until October 2023, which are in JPG or PNG format, we were left with 211,429 unique ad links across all categories, which we scraped using the Selenium package.

After scraping all ads, we used YOLO object detection to identify ads that included cars, which resulted in 1,400 ads. From this set, we manually evaluated the ads to identify those that were ads from automotive brands, which resulted in 543 unique ads and used this as our data set. These 543 unique car ads have a timestamp between September 2015 and October 2022, of which the majority ($> 93\%$) has a timestamp of 2020 or later.

Scraping Flickr

To scrape 1,000 images from Flickr based on the keywords "ruggedness" and "luxury", we utilized a Flickr scraper repository from GitHub (`https://github.com/ultralytics/flickr_scraper`).

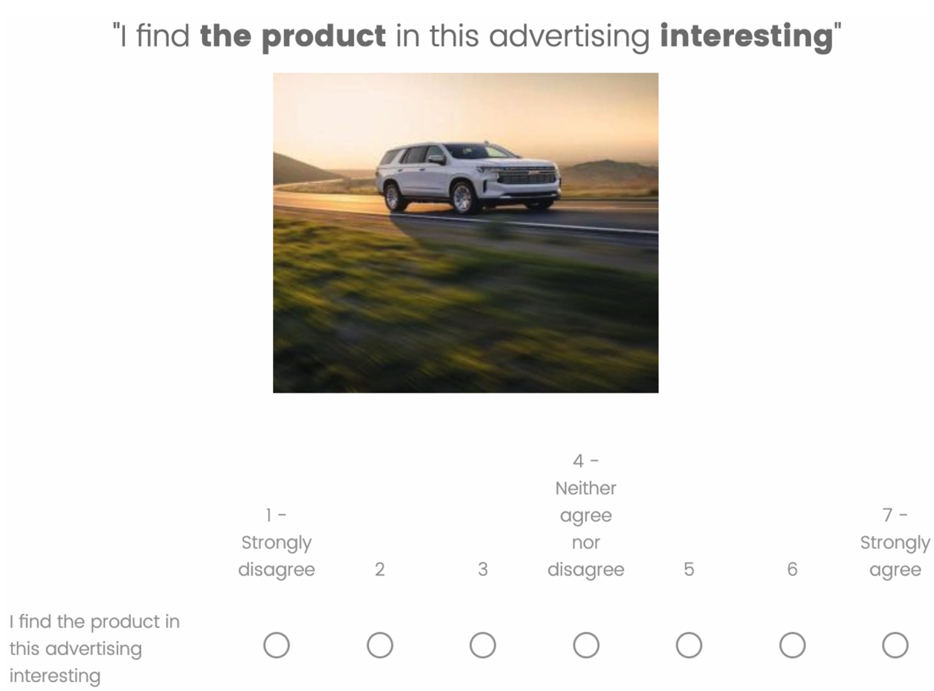
Appendix B: Survey questions

In all studies, we executed a similar survey setup. Each respondent is asked to rate 10 ads on several questions. In study 1 and 2, participants are asked to rate the 10 ads on the purchase funnel metrics. First, each respondent rates a random rotation of 10 ads on attention ("This advertisement would stand out in comparison to other advertisements"), then the same 10 ads in random rotation on interest ("I find the product in this advertising interesting"), desire ("I like the product in this advertisement"), and purchase intent ("If I were in the market for a car right now, I would buy the car in this advertisement"). See an example in Figure 6.

In study 3 and 4 we used the same setup with two additional questions on the perception of ruggedness ("This advertisement looks rugged to me") and luxury ("This advertisement looks luxurious to me"). These two questions are included at the beginning of the survey.

An attention check is included in the middle of all studies ("Check '2' to show you pay attention"). Participants who did not pass the attention check were not allowed to proceed and removed from the study.

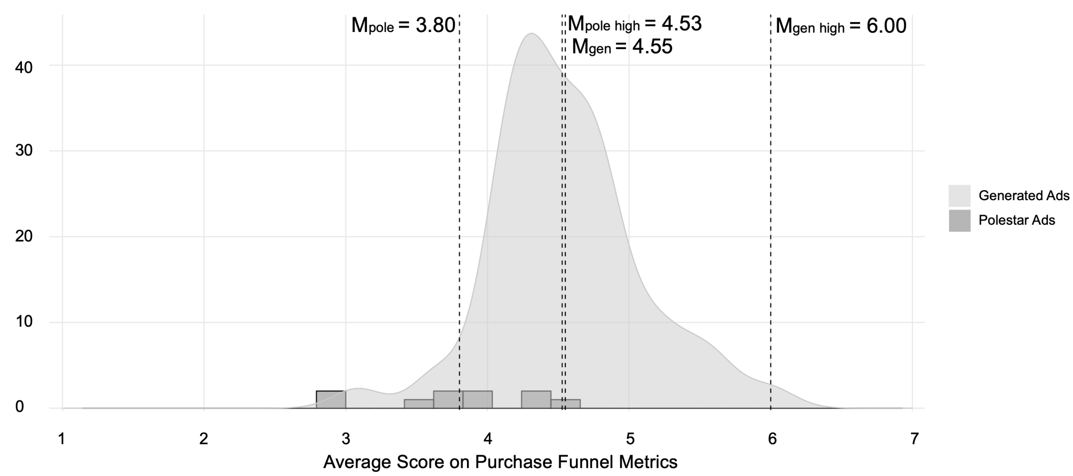
Figure 6: Survey Example Question (Interest)



Appendix C: Distribution of Consumer Ratings of Polestar vs. Generated ads

Figure 7 presents insights in the distribution of the actual Polestar ads and the generated Polestar ads. It is important to note that only 10 different Polestar ads were available. That considered, comparing the average of the generated ads to the actual Polestar ads, we find that the generated ads outperform the actual Polestar ads significantly ($M_{\text{polestar}} = 3.80$, $M_{\text{generated}} = 4.55$, $t = 4.16$, $p < .01$). When we turn to the best performing ads, we find that the best performing actual Polestar ad falls far below the best performing generated ad ($M_{\text{highest polestar}} = 4.53$ vs. $M_{\text{highest generated}} = 6.00$). More specifically, 46% of the generated ads perform better than the best Polestar ad that was actually paid for and distributed.

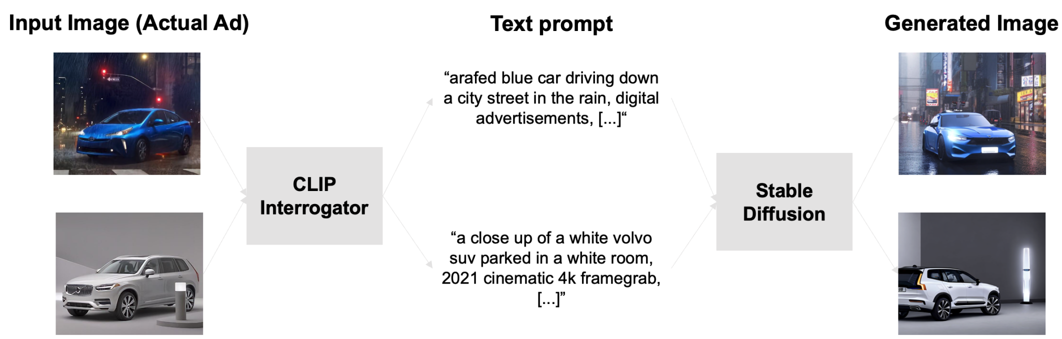
Figure 7: Distribution of Average Consumer Ratings of Polestar and Generated ads



Appendix D: CLIP Interrogator

In Figure 8, we present the workflow used in conjunction with the CLIP Interrogator. We provide the CLIP Interrogator with actual car ads, which are then transformed to a single textual description by CLIP ("Text prompt"). We then use this text as the input for our GDM, which generates a new image based on this text prompt.

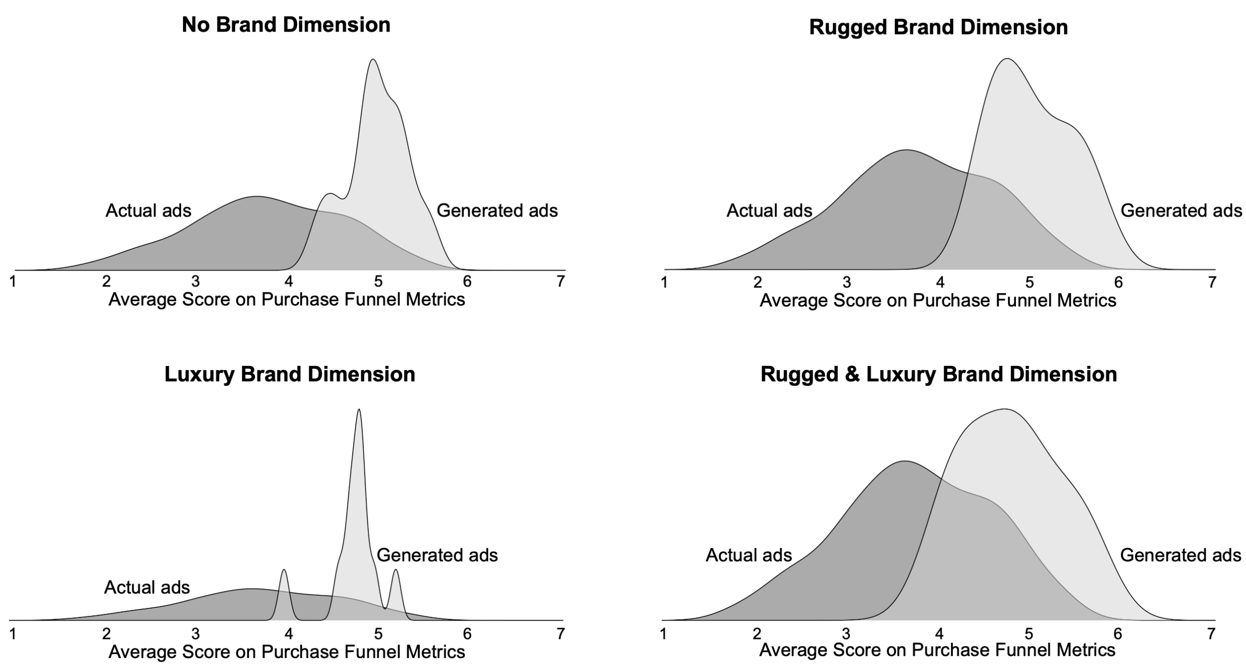
Figure 8: Workflow CLIP Interrogator



Appendix E: Distributions of Actual and Generated ads - Study 4

As in study 1, we examine the distribution of the average ratings across funnel phases of the generated ads including the brand dimensions and compare these to the distributions of the actual competitive ads from this study (see Figure 9). It can be seen that the generated ads, irrespective of the added brand dimension, show similar results compared to the actual competitive ads. All generated ads including no brand dimension, rugged, or luxury brand dimension, or both brand dimensions show similar performance, and all outperform actual competitors' ads ($M_{\text{actual}} = 3.75$, $M_{\text{no dimension}} = 4.94$, $M_{\text{rugged}} = 5.00$, $M_{\text{luxury}} = 4.73$, $M_{\text{rugged \& luxury}} = 4.78$).

Figure 9: Distribution of Average Consumer Ratings of Actual vs. Generated Ads for Individual Brand Dimensions (Study 4)



Appendix F: Comparing Top Generated Ads to Training Ads

Table 7 shows the performance of the best performing generated ads compared to the actual ads which are used as training material. The table shows that the generated ads outperform the actual training ads significantly on all purchase funnel metrics except purchase intent.

Table 7: Comparing Top Generated Ads to Training Ads

Dimension	**Group**	**Descriptives**		**Welch's t-test**
		M	SD	t
Attention	Generated Ads	5.26	1.53	
	Training ads	4.64	1.69	5.65**
Interest	Generated Ads	5.60	1.21	
	Training ads	4.80	1.75	8.29**
Likability	Generated Ads	5.57	1.28	
	Training ads	5.08	1.63	5.08**
Purchase Intent	Generated Ads	5.09	1.67	
	Training Ads	5.34	2.01	6.13**

$^{*}p \leq .05$; $^{**}p \leq .01$.

Appendix G: Fluency of Actual and Generated Ads

We posit that the generated ads might be capable of learning high-level averages that are associated with prototypical advertisements. In this study, we assess if that argument holds by measuring the fluency of all actual ads, and all 130 generated ads from study 1, 3 and 4.

Method

We measure the fluency of the advertisements by asking 1,001 participants (542 females, 454 males, 4 preferred not to state, $\text{age}_{\text{range}} = 25 - 75$), acquired over Prolific (all participants have a driving license and are thus potential customers) to rate 543 actual and 130 generated ads on the level of fluency ("The process of studying this advertisement was…") using a Likert scale ranging from 1 to 7 (1 – Very difficult, 7 – Very easy) (Graf et al. 2019). Each respondent rated 20 images, this resulted in a total of 20,020 individual ratings on 672 images. Each participant rated 20 ads per survey, and all ads are rated 30 times on average.

Results and discussion

We compare the average fluency of all actual and generated ads, the averages of the top 10% and 25%, and the fluency of the highest scoring generated ads compared to actual training images of the model. Table 8 shows that, on average, the 130 generated ads are more fluent than the 543 actual ads ($p < .01$). The top 25% of the generated ads are perceived as more fluent compared to the top 25% of the actual ($p < .01$) and the same holds for the top 10% ($p < .05$). Lastly, comparing the training images of the highest scoring actual ads on all phases to the highest scoring generated ads, we find significant results ($p < .01$). In terms of the underlying distribution, we find that 67% of the generated ads turn out to be more fluent than the average of all actual ads ($M_{\text{actual}} = 4.80$, $M_{\text{generated}} = 5.22$).

Table 8: Fluency of Actual and Generated Ads

Dimension	Group	Descriptives		Welch's t-test
		M	SD	t
100%	Generated Ads	5.22	1.49	
	Actual Ads Dataset	4.80	1.65	15.36**
25%	Generated Ads	5.70	1.24	
	Actual Ads Dataset	5.57	1.28	3.04**
10%	Generated Ads	5.88	1.17	
	Actual Ads Dataset	5.73	1.19	2.20*
Training Set	Generated Ads	5.73	1.21	
	Training Ads	5.13	1.49	9.22**

$^{*}p \leq .05$; $^{**}p \leq .01$.

These observations suggest that GDMs such as Stable Diffusion are capable of generating more fluent ads than what it is trained on. An explanation for this could be that GDMs are trained to create aesthetically pleasing images by averaging various image elements of vast image databases. Such content may appear subjectively more familiar and can be more fluent to process.